European Journal on Criminal Policy and Research
https://doi.org/10.1007/s10610-025-09643-y

Policing *Across* and *Within* Metaverses: How Immersiveness Changes the Role of Police in Cyberspace

Isadora Neroni Rezende[1] · **Emmie Hine**[1,2] · **Mariarosaria Taddeo**[3,4] · **Luciano Floridi**[1,2]

Accepted: 5 August 2025
© The Author(s) 2025

Abstract
Explicit abusive behaviour in the metaverse is increasing, so regulators and other supranational bodies are considering whether this immersive virtual space should be policed and how. In this article, we address this question by arguing that involving national and international police organizations in the metaverse is necessary and demands regulatory action. First, we show that the scope of criminal law should extend to the metaverse and that a policing mechanism is needed to ensure its enforcement. Then, we highlight the shortcomings of existing ways of policing the Web and argue that only a co-governance system involving platforms, online communities, and law enforcement agencies can address the challenges of policing the metaverse. We examine how this undertaking may work using two Levels of Abstraction (LoAs). Through the first LoA, we analyse the metaverse environment from the outside and investigate how policing efforts should be allocated between States across multiple platforms, suggesting ways to solve jurisdictional conflicts. Through the second LoA, we analyse a given metaverse from within and assess the limits of police powers concerning user surveillance and the possibility of performing virtual 'arrests'. In both cases, we show how the immersiveness of the metaverse will change the role of the police online. We conclude the article with recommendations for establishing a policing mechanism in the metaverse.

Keywords International police cooperation · Virtual reality · Metaverse · Policing · Internet policing

✉ Isadora Neroni Rezende
isadora.neroni2@unibo.it

[1] Department of Legal Studies, University of Bologna, Via Zamboni 22, Bologna 40100, IT, Italy

[2] Yale Digital Ethics Center, Yale University, 85 Trumbull Street, New Haven, CT 06511, US

[3] Oxford Internet Institute, University of Oxford, 1. St. Giles', Oxford OX1 3JS, UK

[4] The Alan Turing Institute, British Library, 96 Euston Rd, London NW1 2DB, UK

Published online: 12 September 2025

Introduction

The metaverse is an emerging type of immersive virtual environment (Hine et al., 2024).[1]Explicit abusive behaviour in this type of environment is a growing concern for regulators and other supranational bodies. The European Parliament and European Union Agency for Law Enforcement Cooperation (EUROPOL) commissioned studies on this topic (EUROPOL, 2022; Madiega et al., 2022), and the EU Commission launched a public consultation on virtual worlds, emphasising the need to ensure their safety and security (European Commission, 2023). The International Criminal Police Organization (INTERPOL) is also investigating how to police the metaverse (Cieslak & Gerken, 2023) and released a report on its challenges for law enforcement (INTERPOL, 2024). Users too are raising demands for safety (Zallio & Clarkson, 2022) and have reported being harassed in the metaverse (Basu, 2021; Sales, 2024). How the metaverse may be policed is the topic of this article. Before developing our analysis, some remarks are necessary to clarify and contextualise the exact nature of the problem.

The metaverse is not a unique environment but comprises a plurality of specific 'metaverses', built according to different logics. Many distinguish between centralised-proprietary (also called 'private') and decentralised-distributed (often defined as 'public') metaverses. Corporations (e.g., Meta or VRChat) develop metaverses as centralised platforms where the business model is based on paid services, such as games, and advertisements targeted to users by leveraging their data. Distributed metaverses are instead run by decentralised autonomous organisations (DAOs), owned collectively by their communities (Nabben, 2021, p. 2). They may be inspired by 'crypto-anarchist' and libertarian thinking and rely on distributed ledger technologies like smart contracts to replace traditional hierarchical governance (Atzori, 2017; Santana & Albareda, 2022). While this distinction revolves around the ownership regime of the metaverse, one may argue that the infrastructural model behind a given platform will not be relevant regarding the demand for safety. Therefore, keeping users safe will raise similar issues in centralised (or 'private') and decentralised (or 'public') metaverses, although policing the latter will be more challenging due to their hostility toward State intervention.

In light of this, the distinction between public and private metaverses should not be based on who owns the platform but on how users experience this virtual environment (Floridi, 2022), as further explained in Sect. 5. Therefore, here we focus on immersiveness, a crucial characteristic of the metaverse, which is its added value but may also elicit malicious behavior and expose users to harm.

Scholars and regulators have only begun exploring the ethical and legal implications of the metaverse (Cheong, 2022; Garon, 2022) and still debate whether specific behaviors committed there amount to criminal offenses (Haber, 2024; Lemley & Volokh, 2017). For

[1] There is no agreed definition of what the metaverse is. The European Parliamentary Research Service describes it as 'an immersive and constant virtual 3D world where people interact by means of an avatar to carry out a wide range of activities', see (Madiega et al., 2022, p. 1) The metaverse is supported by extended reality (XR) technologies, a spectrum that includes virtual reality (VR, when users are immersed in a virtual environment, often with a headset), augmented reality (AR, where virtual information is juxtaposed with the physical world), and mixed reality (MR, where objects of the physical and virtual worlds are presented together in a single display) (Milgram & Kishino, 1994) Contemporary versions of the metaverse are distinguished from 'proto-metaverses' like *Second Life*, *Fortnite*, and *Roblox*, which lack the immersive (XR) element.

instance, financial crime is significant in crypto-markets and may soon expand in this environment (Mackenzie, 2022), while live speech and quasi-physical interactions may increase other forms of crime in the metaverse, as we will show in Sect. 2.

Against this background, a study by the European Parliament asks whether we should create a 'metaverse criminal justice system' (Madiega et al., 2022, p. 8). In this article, we address this question by arguing that involving policing organisations in the metaverse is necessary and demands regulatory action. After showing that the metaverse should be policed, we examine how immersiveness will change police activities online to offer policy recommendations using two levels of abstraction (LoAs) (Floridi, 2008). Through the first LoA, we observe the metaverse from the *outside*, analysing how States can ensure an adequate distribution of policing powers across multiple metaverses. Immersive technologies create a new kind of virtual space where avatars will move around and interact as in the analogue world. The consequence will be to prompt responses by public authorities, in terms of real-time assistance to users and crime prevention, as well as transnational police cooperation. Through the second LoA, we examine how policing should work *within* a given metaverse, exploring what powers the police can exercise and under which circumstances. Immersiveness generates new avenues for police action online and, therefore, risks to users' fundamental rights, which regulators should address too.

The metaverse is a pan-jurisdictional space, so it should be governed at the transnational level. Therefore, our analysis offers some policy recommendations addressed to States at the international level but also considers the EU context due to its potential Brussels effect. We do not provide an exhaustive analysis of all the issues related to policing the metaverse, such as using emotional data in surveillance or ensuring cross-border access to digital evidence. These issues are metaverse-*relevant*, but not metaverse-*specific*.

The rest of this article is structured as follows. In Sect. 2, we explain why criminal law should apply to some metaverse behaviours. In Sect. 3, we argue that existing ways of policing the Web are not enough to ensure the enforcement of criminal law and that specific regulatory intervention is needed to address the criminal potential of the metaverse. In the remainder of the analysis, we examine how metaverse policing should be regulated at the two LoAs. In Sect. 4, we analyse how police efforts should be allocated among States and suggest ways to solve jurisdictional conflicts. In Sect. 5, we assess how police powers should be exercised within a given metaverse by focusing on the limits to platform surveillance and the possibility of performing virtual 'arrests' to prevent crimes. In Sect. 6, we conclude the article by suggesting some recommendations for policing the metaverse.

On the Applicability of Criminal Law in the Metaverse

Identifying harmful behaviour and defining crimes are a priority to ensure user safety in virtual worlds (INTERPOL, 2024, p. 12)—even more so, as immersive technologies will also offer new avenues for committing crimes. Experts are debating whether (and which) criminal law should cover metaverse actions. Interpretations vary according to national laws and theoretical approaches (Gómez-Quintero et al., 2024; Haber, 2024). In the past, some have raised 'law of the horse'[2] arguments, contending that illicit actions within the metaverse do not raise new legal questions and should be treated, where possible, as regular

[2] See (Easterbrook, 1996; Lessig, 1999).

crimes (Brenner, 2001; Laue, 2011). Nonetheless, many of these claims were elaborated in reaction to the first developments of virtual worlds (e.g., multi-player games) and thus fail to acknowledge the potential of immersive technologies of the latest versions of the metaverse.

We argue that the scope of criminal law should extend to the metaverse and that, to do so, specific policy action for both existing cybercrimes and emerging criminal behaviours is required to address gaps in legislation. This is because (i) the applicability of criminal law in the metaverse will justify the involvement of State authorities in its policing, in the terms explained in Sects. 4 and 5; (ii) an effective police response, including at the supranational level, will also depend on clear legislation on the matter (INTERPOL, 2024, p. 11).

The distinction between *illicit*, *illegal*, and *criminal* behaviour illustrates why regular cybercrime will become more problematic in the metaverse, and why new virtual behaviours should be criminalised. *Illicit* conduct refers to actions forbidden or disapproved of by moral or social codes (including the law) (Stevenson, 2010). When some conduct directly violates positive law, it is also considered *illegal*. At the same time, illegal behaviour is not always *criminal*. Only actions that wrongfully cause significant harm to others can be criminalised (Ashworth, 2006, p. 31). Indeed, the harm principle is a cornerstone of criminal law in common and civil law systems (Ashworth, 2006; Keiler & Roef, 2019). Criminal law traditionally focuses on 'hard' (i.e., physical) harms, although it has expanded to 'soft' harms of a moral or emotional nature (Brenner, 2008, p. 3), leading to the criminalisation of gambling, harassment, and stalking. However, criminal law does not encompass *any* harmful behaviour, such as trifling wrongs, and presupposes that perpetrators deserve blame for their actions. In fact, although the scope of criminal law protection has expanded, some argue that the legal approach remains selective and fragmented, often failing to address the full spectrum of harm (Michalowski, 2016, pp. 182–185). Traditional legal frameworks tend to focus narrowly on individual wrongdoing, often neglecting the broader structural causes of harm, particularly when they originate from systemic inequalities or powerful institutions (Hillyard & Tombs, 2007, pp. 12–13). Instead, an approach based on social harms would widen the scope of protection to include physical, economic, emotional, psychological and cultural harms — issues that sometimes fall outside the remit of criminal justice despite their significant social impact (Hillyard & Tombs, 2021, pp. 23–24; Pemberton, 2007, p. 37). Therefore, this perspective should be incorporated into discussions about the criminalisation of behaviour in emerging digital environments such as the metaverse. This would ensure that harmful virtual conduct is recognised and addressed appropriately, without being limited by conventional definitions of crime.

In the metaverse, illicit behaviour will generate both soft and hard harms that should be addressed by criminal legislation. In particular, the quality and quantity of harmful experiences for users will increase because of the combination of three features of immersive technologies: (i) interactivity; (ii) telepresence; and (iii) immersiveness (Mütterlein, 2018).

Concerning interactivity, the metaverse increases users' capabilities to influence the form or content of the digital environment, adding the possibility of 'spatial harms' through gestures, postures, and digital assets (World Economic Forum, 2023b). Individuals will be harmed not only by content (e.g., hate speech) but also through interactions that will resemble analogue-world ones. As a three-dimensional space, the metaverse will provide new avenues for offenses like stalking, sexual harassment, assault, or virtual 'rape' (Sethi, 2022, pp. 2–3). From the early days of the metaverse, users (especially those using female avatars) reported distressing invasions of their personal space, often coupled with crude

sexual gestures (Basu, 2021). In 2024, British police started investigating the virtual gang rape of a girl below the age of 16 (Sales, 2024) and different inquiries have highlighted that all kinds of sexual assaults (especially against women's avatars and minors) are a matter of routine in the metaverse (Bates, 2025; Centre for Countering Digital Hate, 2023; SumOfUs, 2022). Although the criminal (or even illegal) nature of such abusive behaviour has been questioned for proto-metaverses (Brenner, 2008), immersive technologies force users to experience these interactions vividly, even without a physical component (Haber, 2024, p. 870), causing genuine psychological harm (Hine et al., 2024). Furthermore, the possibility to move around and interact with virtual objects will also generate new forms of property crime: criminals may steal or damage metaverse assets, or intrude into users' personal space, generating emotional distress or financial loss in the analogue world as well (INTERPOL, 2024, p. 14). This type of crimes dates to the early days of social worlds. In 2005, for example, a Japanese man was arrested on suspicion of 'virtual mugging' for using a bot to rob and beat up characters in the *Lineage II* game (Knight, 2005). A few years later, a Dutch teenager was also arrested for stealing virtual furniture worth €4,000 from the *Habbo Hotel* platform (BBC, 2007). Some countries – such as China, Japan, and the United Kingdom –now recognise the possibility of taking legal action against the deprivation of digital assets, provided that they have monetary value in the analogue world (Fortis, 2023; Qin et al., 2025, p. 4). However, some argue that in the metaverse this protection should extend beyond financial assets to include digital possessions that hold emotional significance for users, even if they lack economic value (Qin et al., 2025, p. 4; Wang, 2023, p. 3).

Metaverse equipment (e.g., headsets) fosters users' sense of telepresence – that is, the state of presence reached using a medium (Mütterlein, 2018, p. 1408) – but also provides opportunities for new forms of manipulation (INTERPOL, 2024, p. 15). Malicious actors can hack these devices and introduce sensory stimuli, altering people's perception of the surroundings (Tseng et al., 2022, p. 4), potentially harming individuals' physical integrity and personal autonomy. One potential avenue for harm is redirected walking, which steers users' physical path by unperceptively rotating the virtual scene. Redirect walking can be used to improve user experience by making users believe they are walking a long distance in a straight line when they are actually walking around a confined space (De Back et al., 2024); however, studies have shown that criminals may employ it to harass users (Casey et al., 2021), causing emotional distress, or exposing them to dangerous situations, for example, by pushing them towards stairs or a street (Gómez-Quintero et al., 2024, p. 11; Nichols, 2022; Wang et al., 2022, p. 22). Equipment such as headsets, haptics, or teledildonics may also offer more realistic sexual experiences and encourage the commission of related offenses, such as the provision of AI-generated child sexual abuse materials (Gómez-Quintero et al., 2024, p. 18). In South Korea, for example, a man was sentenced to four years imprisonment for producing and storing sexually explicit contents involving minors he groomed in the metaverse (Adejumo, 2022).

Combined with interactivity and telepresence, emotional immersiveness will also increase the harmfulness of existing crimes and other illicit behaviours in the metaverse. Research shows that immersive technologies increase users' psychological presence or their sense of 'being there' (Makransky & Mayer, 2022; Murray & Sixsmith, 1999); that is, a feeling of embodiment created by freedom of movement; stereoscopy, a broader field of view that strengthen the impression of being within an environment (Cummings & Bailenson, 2016); and a sensory realism from visual, auditory, tactile, and even olfactory stimuli (Cadet

& Chainay, 2020; Gromer et al., 2019). Such features create a correlation between users' virtual bodily actions and their physical counterparts (Cummings & Bailenson, 2016, p. 3). As a result, users will experience metaverse events and encounters almost like their analogue correspondents (Dwivedi et al., 2022, p. 4). Research shows that virtual reality (VR) experiences trigger many of the same psychological and physiological reactions as experiences in the physical world (Martens et al., 2019). Therefore, illicit behaviour committed in the metaverse may have illegal consequences that are best addressed with criminal law instruments (Haber, 2024, p. 845). Illustrating this, police in the UK are investigating a case of 'virtual rape' that reportedly left a girl traumatised, even if it is unclear whether existing criminal law will be able to address her case (Farrant, 2024). Thus, there is a chance that illegal actions may remain unpunished. The risk of harm is even more serious for children, who may be more vulnerable to exploitation and struggle to distinguish what is real and what is make-believe (Sales, 2024). In *Horizon Worlds*, it has been found that nearly 7 out of 10 worlds are visited by minors, including 'Mature Worlds', where Meta permits sexually explicit content, legal drugs promotion, and gambling (Centre for Countering Digital Hate, 2023, pp. 4–5). Although adult users are permitted access to such environments, there are no safety measures in place to prevent children from entering. Consequently, researchers have uncovered many instances of harassment against underage users, who were frequently subjected to sexually explicit insults (Centre for Countering Digital Hate, 2023, p. 6).

In this perspective, existing crimes aimed at causing fear and emotional distress – such as blackmail, cyberbullying, incitement to self-harm, harassment, or stalking – will have a heightened negative impact on users, as immersion will increase the sense of threat perceived by victims. Similar effects will be achieved by other cybercrimes, like identity and data theft, illegal use of deepfakes, or property crimes, which usually cause non-material damages beyond financial loss. Griefing and other violations of game rules may not be considered trifling wrongs anymore, as immersion will enhance their emotional and long-term impact on victims (Oppenheim, 2022). Beyond individual harms, the metaverse may also be leveraged as a tool to commit crimes against public safety, like propaganda, misinformation, radicalisation, and indoctrination, also related to terrorism. For instance, the Australian Federal Police has warned that extremist groups are infiltrating proto-metaverses such as *Roblox* to target child users (Australian Associated Press, 2023). In these cases, immersion may intensify the potential for manipulation, causing users to make uninformed decisions with harmful consequences at the collective level (INTERPOL, 2024, p. 15). Terrorist groups may even use metaverse spatiality to rehearse their plans in three-dimensional environments, leading to improved coordination of real-world attacks (Hine et al., 2024; INTERPOL, 2024, p. 15). Exemplifying this, a self-radicalised Singaporean teenager was issued with a restriction order for joining ISIS-themed settings on *Roblox*, where he role-played as a soldier of the Caliphate in replicas of conflict zones such as Syria and Marawi City. The teenager also used footage from the game to create propaganda videos depicting ISIS factions carrying out attacks (Singapore Ministry of Home Affairs, 2023). Another controversial case involves a Russian teenager who was sentenced to five years in prison for allegedly planning to blow up a virtual Federal Security Service (FSB) building in *Minecraft* (France-Presse, 2022).

Empirical evidence of crimes committed in the metaverse is, often, still limited to specific incidents reported in the news, so more extensive research would certainly be needed to gain a comprehensive understanding of the phenomenon. Nevertheless, the accounts and

scenarios presented show the risks of the metaverse causing actual harm to people, and yet, users may not be protected under existing criminal legislation, which has been mainly designed to address the physical and emotional harms of the analogue world. A theoretical and a practical reason supports the need for enacting specific legislation on metacrimes (i.e., crimes committed in the metaverse).

At the theoretical level, harms prompted in the metaverse will differ in nature and intensity, so specific legislation is needed to criminalise such behaviours or adjust sanctions accordingly. Concerning the nature of the harms, traditional crimes, such as rape or kidnapping, may be simulated in the metaverse, but the physical element of the *actus reus* will be missing. Therefore, these experiences will not compare to their analogue correspondents, and the harm suffered by victims will be qualitatively different. If the criminalisation of rape and kidnapping aims to protect values such as (physical) sexual integrity and personal freedom, corresponding metacrimes will have to be mainly construed to address emotional distress. Speaking instead of the intensity of the harm, the negative consequences of existing cybercrimes (e.g., hate speech or indoctrination) will easily extend to the metaverse but may deserve more severe sanctions when committed there, considering the higher intensity of the psychological impact on users.

Enacting ad-hoc legislation for crimes committed in the metaverse is also needed for a practical reason: to ensure the effectiveness of transnational criminal cooperation. Following the 'law of the horse' argument, one may contend that existing legislation may be enough to address illegal behaviours since, for example, virtual rapes will not qualify as such under national law but may still be prosecuted as battery or harassment. However, such an approach will likely cause fragmentation because States may not agree on the legal qualification of the same conduct or have conflicting definitions for similar crimes committed in the metaverse, as in the case of the 'virtual rape' mentioned above. Under such conditions, States may refuse to cooperate and bring suspects and accused to justice, undermining the effective enforcement of criminal law for crimes perpetrated in this kind of environment.

The need to apply criminal law in the metaverse highlights the importance of policing virtual worlds. Platforms and the police are already active on the Web, but their efforts may not be enough to address the challenges of immersive environments. The following section explains why and provides arguments for establishing a specific policing mechanism for the metaverse.

From Platform Moderation To Policing the Metaverse

If criminal law applies in the metaverse, then policing is necessary to ensure its effective enforcement. In this section, we examine how private actors and the police act on the Web to moderate user behaviour, highlighting the strengths and weaknesses of current approaches. Privately enforced mechanisms of content moderation are prominent but arguably problematic from a fundamental rights perspective, while police initiatives are scarce but potentially more safeguarding for users. Given this situation, we argue that only the combined intervention of private and public entities can effectively address crime threats in the metaverse.

In the traditional Web, platforms address user behaviour through Terms of Service (ToS) and/or community guidelines. Platforms manage infringements through content moderation mechanisms (Hine, 2023), but, occasionally, they also determine criminal liability under

national legislation (Toolkit: Cross-Border Content Moderation, 2021). Content moderation can be oriented more towards platform moderators, like on Reddit, or automated systems can be used as a first line of defence, such as on Facebook. Many platforms use some combination of both.

Online communities also have a role in ensuring public safety and order on the Internet. Traditionally, community self-policing has contributed to the police function in society (Robinson & Scaglion, 1987, p. 116), and it has applicability in virtual environments as well (Headleand et al., 2020, p. 121). For example, in *World of Warcraft*, game communities established regulations to counter misconduct, such as cheating, toxic behaviour, and verbal aggressions, and participate in self-surveillance to enforce them (Collister, 2014). In Meta's *Horizon Worlds*, 'volunteer moderators' monitor spaces in real-time and intervene in escalating situations by muting or removing other users in more serious cases (Meta, 2024).

Although platform content moderation systems and communities will be helpful in the metaverse to manage user behaviour, more needs to be done to keep users safe, as current mechanisms are not enough to address the crime potential of the metaverse. Two reasons support this need. The first relates to the quality of the policing from a user fundamental rights perspective. Even when no crime occurs, policing affects how individuals experience an environment, whether analogue or digital. Generally, police efforts increasingly focus on offenses yet to be committed. Their presence in specific areas is often linked to deterrence (Marthews & Tucker, 2017, pp. 438–439) but may also affect communities of colour and other disadvantaged groups disproportionately, fostering mistrust between communities and the police (Babe Howell, 2016; European Roma Rights Centre, 2023; Jackson et al., 2023). Police surveillance and abuses may have a chilling effect on how users enjoy the metaverse and express themselves in this virtual space. Moreover, upholding fundamental rights standards in police conduct is essential for the consequences that irregularities may further have on criminal proceedings in the analogue world. It is part of a general debate on digital sovereignty (Floridi, 2020). Those who commit offenses in the metaverse should eventually bear the consequences of their endeavours in one legal system, for example, by paying fines, serving prison time, or another alternative punishment. Consequently, the prevention and ascertainment of criminal acts in the metaverse must be governed by similar guarantees, for instance, due process considerations such as equality, accountability, and the right to an effective remedy.

Given how policing interferes with individuals' rights, those responsible for this task in the metaverse should comply with the highest standards of independence and impartiality. It is doubtful, however, that platforms and online communities acting alone will meet such requirements, as highlighted by the shortcomings of current content moderation mechanisms. These often lack oversight mechanisms or checks and balances provided by law, meaning that private platforms retain wide latitude on how they handle content (European Commission, 2020, p. 25). Studies have also shown that, when claims about the illegality of content are uncertain, platforms will often remove the information to avoid liability (Riis & Schwemer, 2018, p. 13; Urban et al., 2017, p. 41), suggesting that businesses do not manage content moderation impartially but according to corporate logics (Qiwei et al., 2024). In addition, there is often a lack of transparency and accountability in the decisions taken unilaterally by platforms. For instance, the European Commission's impact assessment on the proposal for the Digital Service Act (DSA) reports that very few online platforms are used to subject their content moderation policies to systematic independent oversight (European

Commission, 2020, pp. 18–19). The same document mentions a Eurobarometer survey, in which 17% of respondents declared that online platforms erroneously removed their content but were never informed about the reason for the removal (European Commission, 2020, p. 26).

In the metaverse, replicating these mechanisms could be very damaging for users. While metaverse platforms currently are using primarily human moderation (Gray et al., 2024; Schulenberg et al., 2023), they could use automated technologies to control avatar behaviour rather than content, for example, by pre-emptively blocking their actions whenever a situation of potential danger arises. Additionally, these tools may be tuned to avoid platform liability, thus resulting in overly stringent restrictions on user interactions (we will elaborate on this in Sect. 5). For this reason, such sensitive decisions should not be entirely delegated to opaque technologies, devised according to platforms' corporate needs.

Similar concerns apply to policing by online communities, which might engage in in-game harassment and bullying (i.e., griefing), phenomena that may disproportionately affect women, LGBTQ+players, and players of colour in VR platforms (Ajibola, 2022; Beres et al., 2021, p. 1; Centre for Countering Digital Hate, 2021; Crawford & Smith, 2022; Outlaw, 2018). Beyond discrimination, accountability is also an issue in online community policing, which often lacks redress mechanisms for affected users. Empirical research indicates that punishment before community 'game tribunals' (e.g., in *League of Legends*) can be too hasty. Behavioural standards of misbehaviour are unclear, and the mere fact of being reported often leads to a sanction (Ehrett, 2016). The users participating in the sentencing appreciate the flexibility of this process, which, however, raises issues regarding the rights to equality and effective remedy. In *VRChat*, a 'social VR' platform, users organised themselves into a vigilante police department. Though it began as a roleplaying game of questioning and 'arresting' willing avatars, 'officers' have also been accused of planting evidence and harassing users (Ajibola, 2022). Against this background, community policing online appears helpful in addressing minor misconduct but insufficient to protect users from abuse.

The second reason for supporting a different policing scheme in the metaverse stems from observing the limited effectiveness of mechanisms enforced only by private actors. When content moderation depends only on platforms' voluntary actions, detection of illegal content, assistance to victims, and coordination with law enforcement are affected negatively. For example, the European Commission has facilitated the adoption of self-regulatory mechanisms in the EU, but these have shown structural limitations in scope and scalability. Given their voluntary nature, these initiatives were limited to their signatories, and compliance could not be supervised or sanctioned appropriately (European Commission, 2020, p. 30). The European Commission's impact assessment on the proposal for a Regulation to prevent and combat child sexual abuse also highlights that many service providers have not implemented measures to detect child sexual abuse materials (CSAM) despite their proliferation online (European Commission, 2022, p. 38). Even when platforms undertake similar initiatives, they often do so unsuccessfully. For example, in the past, Twitter (now X) failed in essential enforcement against CSAM by allowing already-flagged images to circulate on the platform (Corse, 2023), while Instagram facilitated a commercial network of CSAM (Horwitz & Blunt, 2023). Voluntary cooperation between private and public actors depends on ad-hoc temporary solutions that are rarely effective in addressing child sexual abuse (European Commission, 2022, p. 49). The DSA now foresees obligations for service

providers in the EU to detect illegal content online. Still, after its adoption, some large platforms have engaged in legal disputes to escape its application (Goujard, 2024), displaying an unwillingness to bear the administrative costs stemming from this legislation.

As we mentioned at the beginning of this section, law enforcement authorities also struggle to tackle criminal behaviour online. EUROPOL's report on the metaverse mentions that only a few countries have made notable efforts to build a presence on social media, although this work will be very different from policing metaverses (EUROPOL, 2022, pp. 24–25). In this context, officers should not just be approachable as on social media, but proactively intervene, for instance, if a user is in danger. In this regard, some pilot projects have begun experimenting with how the police should act in the metaverse and proto-metaverses. In France, for example, the NGO *Enfant Bleu* partnered with *Fortnite* to launch an undercover police operation that involved over 1,200 children and teenagers who reported personal issues and situations of danger (Agence France-Presse, 2023). While successful, such initiatives are sporadic, time-bound, and insufficient to address crime in immersive environments. In fact, police forces will likely struggle to keep up with the task of online policing, due to the structural lack of knowledge and technical capabilities to inspect complex digital services (European Commission, 2020, p. 22). Moreover, the legal requirements and practical hurdles of police's international cooperation represent an additional concern for future metaverse policing. Currently, cooperation is often limited to high-profile cases and requires burdensome interactions between national authorities (Greenberg, 2022), while in the metaverse, crimes that generally do not trigger such mechanisms (e.g., stalking) will also demand a prompt response by the police.

Both the limitations concerning private-led initiatives and law-enforcement efforts to keep the metaverse safe can be overcome if policing the metaverse is framed as a task for private and public entities acting together in a co-governance regime. While providers and communities may be best placed to detect criminal behaviour on their platforms, law enforcement will need to supervise their actions to check compliance with fundamental rights, be able to intervene directly, and ensure follow-up in the analogue world through international cooperation. Such coordination is necessary because individuals cannot be prosecuted, convicted, or imprisoned in the metaverse, a virtual space where physically coercive acts (e.g., arrests) cannot be performed. Restraining individuals and imposing criminal sanctions will thus remain an exclusive competence of criminal justice authorities in the analogue world. Instead, policing activities that do not require physical force – such as crime prevention, assistance to the public, and surveillance – will be possible, and should be ensured, in the metaverse. These tasks could be performed or facilitated by a dedicated supranational body, structured in the terms outlined in Sect. 4, ensuring platform compliance and coordinating jurisdictional conflicts when crimes are detected. As argued in Sect. 2, specific legislation should be enacted to set up this policing mechanism. In the EU, the DSA already provides a legal basis for tackling illegal content online. Still, this legislation would arguably leave out the new forms of criminal behaviour that, as described above, are now made possible by spatial harms. An ad-hoc framework is thus necessary to legitimise police action in the metaverse, avoid legal uncertainty for private entities, and foster coordination with public authorities. The new regime will also have to allocate policing efforts among all the actors involved, as an unorganised police presence could cause excessive surveillance or create jurisdictional conflicts between States. The following section explores policy options to organise police cooperation in the metaverse to address these issues.

Policing Across Metaverses: Transnational Cooperation in the Metaverse

Having established the criminal relevance of some metaverse behaviours (Sect. 2) and the need for a specific mechanism for metaverse policing (Sect. 3), we now address how this may work in practice. We do so through a LoA that examines how policing efforts should be allocated across multiple metaverses. By reviewing different models of international police cooperation, we explain how a supranational body may perform this task in the metaverse, and how jurisdictional problems can be addressed.

Historically, international police cooperation has been structured according to two models. On the one hand, police forces may engage in bilateral or multilateral agreements on extradition or information sharing. That is the case of INTERPOL and EUROPOL, which do not have authoritative powers but rely on those exercised by national police forces. On the other hand, States may cooperate by establishing a supranational police body, endowed with autonomous powers to arrest and collect evidence. However, because this solution entails a significant waiver of sovereignty, States have been reluctant to pursue this path (Deflem, 2000, pp. 749–750), with limited exceptions. In the EU, for example, the European Anti-Fraud Office (OLAF) conducts administrative investigations on fraud and other crimes against the financial interests of the Union, performing acts like on-the-spot checks and inspections. The European Public Prosecutor Office (EPPO) operates in the same domain and prosecutes individuals before Member States' domestic courts. Both Offices have autonomous and enforceable powers in national jurisdictions, but they mainly act *ex-post*, i.e., after the criminal facts have occurred. At the national level, federal states also implement forms of police cooperation across their federated territories. For example, the United States Federal Bureau of Investigation (FBI) investigates federal crime and performs preventive surveillance and intelligence operations.

Translating these two models to the metaverse, two forms of police cooperation are foreseeable: *intergovernmental coordination* and the creation of a *supranational police body*.

In intergovernmental coordination, national law enforcement agencies (LEAs) and platforms would police the metaverse while being coordinated by a supranational agency like INTERPOL or EUROPOL. Such an entity would facilitate coordination if crimes were detected, ensuring that information is transmitted to the competent analogue-world authorities, but it would not have the power to perform coercive acts. Arguably, this model would be easier to set up at the political level but would not ensure an equal allocation of State efforts, since national LEAs would still perform policing. For the same reason, 'positive' conflicts of jurisdiction may arise between States imposing their competence over the same portion of the metaverse, while 'negative' conflicts would appear when no police force claims jurisdiction over a given space, leaving users unsafe.

Conversely, an optimal—but maybe utopian—solution would be creating a supranational body ('the metaverse police') endowed with autonomous powers to police the metaverse. This entity would be composed of detached officers from national LEAs, following the model of the United Nations Blue Helmets (Cox, 1999). If this seems unorthodox, one should consider that several armies in the world, for example, have reorganised themselves to add a new branch entirely dedicated to defensive and offensive operations in cyberspace, besides the conventional domains of war: land, sea, air, and space. Such a development could also affect law enforcement, thus supporting the creation of a supranational body to address

metaverse crimes. This scenario, where a unique entity performs policing, would mitigate jurisdictional issues and ensure an ideal distribution of policing efforts across metaverses.

As explained in Sect. 3, an ad-hoc legal basis would be necessary to set up one of these policing mechanisms. An international treaty, gathering the support of as many States as possible, would be the ideal framework. However, this may not be politically possible in the early phases of metaverse development. Conversely, undertaking this initiative at the macro-regional level, like in the EU, appears more feasible and could still impact many users thanks to the so-called Brussels Effect (Floridi, 2021, p. 217). Besides, the history of the EPPO shows that 'starting small' can be a promising path in this domain: conceived as a form of 'enhanced cooperation' among a restricted group of Member States, the Office now operates in 24 EU countries. Should a similar initiative be taken for the metaverse, the policing mechanism could expand to other States beyond the EU. Currently, the Treaty of the Functioning European Union (TFEU) allows adopting, respectively, measures to strengthen the harmonisation of the internal market (Article 114 TFUE) and police cooperation (Article 87 TFEU). Based on these legal provisions, an EU regulation would be the appropriate framework to create a policing mechanism in the metaverse, establishing common rules for digital platforms and a mechanism of cross-border cooperation between LEAs in the EU.

Nonetheless, enacting an international instrument for police cooperation would not be helpful if the States involved did not have a robust jurisdictional claim over the metaverse platforms involved. As underlined by INTERPOL, a major challenge this new environment presents to law enforcement is determining national jurisdiction (INTERPOL, 2024, p. 20). In fact, the police function has always been tied to the principle of sovereignty, which allows States to exercise public powers within their territories. However, the metaverse is a fully digitised environment where no State can claim territorial sovereignty. Moreover, before criminal acts are committed, determining national jurisdiction over a portion of the metaverse is even more challenging, given that the nationality of the perpetrator and the victim are still unknown. Establishing the competence of a police body in a preventive context, in a fully transnational environment like the metaverse, is thus a relatively new issue for criminal justice. Only the high seas or outer space offer similar examples of non-territorial spaces in the analogue world. However, these do not pose the same crime risks to the average individual as the metaverse. The problem of determining jurisdiction also emerges *ex-post facto*, when multiple States can claim jurisdiction over the same metaverse crimes.

The most appropriate solution for the metaverse should be found in the criteria for establishing State jurisdiction in international law (Oxman, 2007; Schmitt, 2017):

- *Territory.* Each State has jurisdiction over the persons, properties, and activities located within its borders;
- *Nationality* is in the form of both *active* and *passive personality* principles. Under the former, the State applies its rules to its nationals and companies established in its territory, even if the crime occurs abroad. The latter anchors jurisdiction to the nationality of the victim;
- *Residence or domicile*. This implies that States can enforce their norms on subjects who reside or domicile in their territory;
- *Protective principle.* It allows States to claim jurisdiction over activities that occur abroad but encroach on their governmental functions (e.g., counterfeiting their currency);

- *Universal jurisdiction*. It defines the right of a State to prosecute some internationally defined crimes (e.g., genocide), even if committed abroad and by foreigners.

Among these, neither the protective nor the universal criteria seem to adapt to the nature and seriousness of metaverse crimes. The lack of a physical dimension in metaverse behaviours makes it impossible to commit serious international offenses, such as war crimes or crimes against humanity, as defined by the Statute of the International Criminal Court. The protective principle could apply to the counterfeiting of digital currencies or official documents, but the specificity of the object of these offenses makes this criterion challenging to extend on a large scale. Consequently, the nationality and residence/domicile principles will be the most useful in determining state jurisdiction in the metaverse, although they have different relevance in *ex-ante* and *ex-post facto* situations.

In the *ex-ante facto* context, the nationality or location of future suspects and victims is still unknown, so that the metaverse police could be competent under another nationality criterion, such as the place of establishment of the platform concerned. This is a recurrent mechanism in EU legislation applying to digital providers, notably in the General Data Protection Regulation (GDPR).[3] Often, EU law also requires that platforms that are not established in the Union, but provide services to many users, designate a representative in its territory.[4] Under this system, digital platforms would fall under the competence of the metaverse police whenever they are established or operate in a State member of the EU or the international treaty mentioned above.

In an *ex-post facto* situation, jurisdictional criteria identify the competent authorities in the analogue world for investigating, prosecuting, and sentencing metaverse crimes. The most adequate basis for jurisdiction should be chosen to ensure the efficiency of the cooperation mechanism, given that the metaverse police, since its first intervention, will have to identify which national LEA must be informed of the alleged criminal facts. Considering this, although potentially applicable, the nationality criterion may not be the best solution. Relying on platforms' place of establishment may create a disproportionate workload for some States. At the same time, claiming jurisdiction for the victim's nationality might be controversial,[5] while that of the suspect may not always be useful because the person concerned might be located outside of their State of nationality.

Conversely, a practical solution is to use the residence or domicile criterion. LEAs may not be willing to initiate burdensome surrender procedures if suspects live in remote countries and are suspected of minor offenses (Lemley & Volokh, 2017, pp. 1071–1073). Hence, physical surrender should be avoided to prioritise information exchange between police forces, through the coordination of the metaverse police. A national LEA would thus be

[3] Art. 3(1) of the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L 119, 4.5.2016, p. 1–88.

[4] Art. 27 GDPR; art. 7 Regulation (EU) 2023/1543 of the European Parliament and of the Council of 12 July 2023 on European Production Orders and European Preservation Orders for electronic evidence in criminal proceedings and for the execution of custodial sentences following criminal proceedings, OJ L 191, 28.7.2023, p. 118–180.

[5] The passive nationality principle has been criticized because it may intrude on the right of another State to prescribe and enforce legislation in its territory or to its nationals, and on the rights of individuals who cannot be expected to know the criminal laws of every State when travelling abroad (Oxman, 2007, § 34).

informed that someone in their territory has committed an offense in the metaverse, allowing them to take the most appropriate course of action. Still, one must consider that if the State where the suspect is located is not a party to the treaty or the EU – continuing with our example – it may refuse to initiate any criminal proceedings. In this case, other States may claim jurisdiction under the remaining criteria (i.e., nationality of the platform, suspect, or victim), even though they would still need the cooperation of the residence country to inform the suspect of their rights in the proceedings, or ultimately to enforce a sanction.

Therefore, as in the analogue world, the ultimate enforcement of criminal law in the metaverse will rest on the States' goodwill. If cooperation is refused, some crimes committed in the metaverse could be left unpunished in the analogue world, but other instruments would still be available in the metaverse to keep users safe, at least temporarily. In this respect, the following section examines under which conditions the police could intervene online against users engaging in criminal behaviour.

Policing Within a Metaverse: Surveillance and Arrests

In this section, we outline further how metaverse policing should function through the second identified LoA, which aims to determine what powers the police can exercise in the metaverse. As anticipated in Sect. 3, police efforts online are mainly devoted to detecting and preventing crimes. In the metaverse, the police will first have to carry out general surveillance activities. Occasionally, the police will also have to interact with avatars in real-time to block criminal offenses. In the analogue world, the instrument dedicated to this task is the flagrancy arrest, but its applicability in a virtual environment is doubtful.

To understand if and how the police could exercise these powers, we analyse how privacy norms limit general police surveillance in public and private places of the metaverse, and if the notion of arrest can be translated into immersive environments. While measures like freezing or blocking users cannot be compared to arrests, they have similar fundamental rights implications. Consequently, we identify the basic legal requirements of these virtual measures by looking at the human rights framework for restrictions of liberty and personal security.

For starters, avatars in the metaverse will be able to move around and will have expectations of where they can find the police or be arrested, and when their activities can be monitored also through data collection. As we wrote in the introduction, the distinction between private and public metaverses should not be founded on who owns the platform when it comes to policing, but on how users experience this environment. As individuals feel immersed in the metaverse, they will perceive the distinction between public and private places as in the physical world; that is, looking at the accessibility of a given location and not its ownership regime (Altman & Zube, 1989; Ruppert, 2006). For instance, Seoul has launched a metaverse allowing users to access government services in a virtual replica of the city (De Almeida, 2023; Maresca, 2023). While this metaverse is built as a centralised platform, its spaces and services are freely accessible to everyone, meaning this environment should be considered a public place. Likewise, Meta defines the metaverse as a 'public place' (*Meta Horizon Worlds Frequently Asked Question| Quest Help| Meta Store*, n.d.) but this may not apply to every space in the platform. For example, the metaverse may include places that should be considered private, like those created for training and work (e.g., Meta

Horizon's Workrooms), or by users to retreat in their bubbles (Falchuk et al., 2018, p. 56). In this perspective, policing in the metaverse must respect the distinction between public and private spaces, as in the physical world, in the terms explained below.

Moreover, policing may be affected by similar limitations as those of the analogue world, meaning that not all surveillance activities will be equally intrusive. While 'beat policing' and other forms of manual observation in the metaverse will be resource-constrained and provide a limited amount of information, real-time acquisition and automated analysis of data will give a more comprehensive picture of users' activities beyond officers' standard capacities of observation (Hine et al., 2024).

These activities will interfere with the right to privacy, which is enshrined in several human rights instruments (Article 12 of the Universal Declaration of Human Rights, UDHR; Article 17 of the International Covenant on Civil and Political Rights, ICCP; Article 8 of the European Convention on Human Rights, ECHR). These provisions protect not only the 'home' strictly speaking, but also business premises, as clarified by international jurisprudence.[6] Moreover, case law generally recognises that individuals may claim some expectations of privacy also in public places,[7] as these serve social and political functions (Cohen, 2013, p. 1927; Habermas, 1991; Regan, 2015). Therefore, police activities should be regulated and supervised in these spaces too.

In the metaverse, the police should be able to access public venues freely and observe avatars' activities 'manually'. However, privacy expectations would require specific safeguards for more intrusive forms of surveillance, like real-time data analysis of user data.[8] Even if the monitoring occurs publicly, such measures should be allowed only for the most serious offenses and require authorisation (or *ex-post* validation) from an independent authority. A similar regime should also apply to policing and surveillance of private metaverse spaces, such as users' private bubbles[9] or those created for work purposes. Specifically, the police would have access to these private venues only in limited circumstances and with previous authorisation by an independent entity, while emergency interventions would be justified only upon a user or platform alert, and if validated *ex-post* by the same authority.

During surveillance, the police may also spot users committing crimes and take the initiative to stop them. Intuitively, these interventions may resemble arrests, but it is unclear whether this is their right legal qualification. Indeed, the etymology of 'arrest' involves the idea of stoppage or staying something ("Arrest", n.d.), but in criminal procedural systems, arrests imply an act of material apprehension (Garner, 2014). In proto-metaverses like *Second Life*, the platform simulated arrests by 'freezing' other avatars and programmatically

[6] In the European domain, the leading case is ECtHR, *Niemietz v. Germany*, judgment of 16 December 1992, App. no. 13,710/88. At the international level, the concept of 'home' in Article 17 ICCP includes the place where the individual 'carries out his usual occupation', as clarified in the CCPR General Comment no. 16 (1988), § 5.

[7] In the US context, see e.g., *Carpenter v. United States*, 585 U.S. (2018). In the European one, see ECtHR, *P.G. and J. H. v. United Kingdom*, judgment of 25 September 2001, App. no. 44,787/98, § 56; ECtHR, *Peck v. the United Kingdom*, § 57; ECtHR, 5 September 2017, *Bărbulescu v. Romania*, App. no. 61,496/08, § 70; ECtHR, *Boulois v. Luxembourg*, judgment of 3 April 2012, App no. 37,575/04, § 63; ECtHR, *National Federation of Sportspersons' Associations and Unions (Fnass) and Others v. France*, judgment of 18 January 2018, App. nos. 48,151/11 and 77,769/13, § 151; ECtHR, 17 October 2019, *López Ribalda and Others v. Spain*, App. nos. 1874/13 and 8567/13, § 89.

[8] Cf., *mutatis mutandis*, USSC, *United States v. Jones*, p. 3, Sotomayor concurring.

[9] In the metaverse, private bubbles may refer to users' homes bases or other portions of space, surrounding the avatar, 'walled off' temporarily for personal use (Falchuk et al., 2018, p. 58).

confining them to a jail cell for some time (Leenes, 2008, p. 107). Nonetheless, these temporary bans had minor consequences on the well-being of users, who could always log off for a period to escape the consequences of the block. Nowadays, similar actions are possible in the new versions of the metaverse (Ajibola, 2022), but since these do not entail a forcible seizure of an individual, they cannot qualify as arrests in strict legal terms. We will instead refer to them as 'blocking measures'.

To be clear, the analysis of the legal requirements for these blocking measures does not concern the arrests that may be performed subsequently in the physical world. There is not necessarily correspondence between virtual measures and physical arrests. Nonetheless, as blocking measures will have implications in the analogue world, they should be governed by similar principles, such as fairness and proportionality. In this sense, some illicit actions, such as vulgar behaviour, will not legitimise any kind of arrest and should instead be addressed by game rules (for instance, through temporary bans); some illegal or criminal behaviours may justify only a blocking measure, while more serious ones will also demand a physical arrest. Furthermore, applying a virtual or physical arrest will not necessarily lead to a prison sentence in the analogue world, such as when the accused is eventually acquitted or sanctioned with a fine or other alternative punishment.

The resemblance between physical arrests and their virtual correspondents should not be underestimated, at least from a fundamental rights perspective. Two reasons support this point. First, if one considers the etymological idea of arrest, blocking measures are also something that stops the perpetrator from interacting normally with the virtual environment. Therefore, even without any physical apprehension, these measures will still have similar psychological and social implications as traditional arrests, also considering that immersiveness will amplify the humiliating and distressful impact for users. Second, as described in Sect. 2, sudden intrusions on avatars' movements can also have physical consequences for users. Even if they do not entail an actual restriction of liberty, blocking measures can be traumatic for individuals because being frozen could generate cybersickness or aggressive responses from users experiencing a loss of control over their surroundings (Behr et al., 2005, p. 671).

Considering these similarities, the fundamental rights regime for arrests should apply to blocking measures in the metaverse. The basic legal requirements for restrictions to individual liberty, like arrests, are set in human rights treaties that protect the rights to liberty and personal security. Article 3 UDHR and Article 9 ICCP provide that '[n]o one shall be subjected to arbitrary arrest and detention'.[10] Under Articles 9 ICCP and 5 ECHR, arrests must always have a basis in the law, and the individuals apprehended shall be informed promptly of the reasons behind the arrest, be brought within a reasonable timeframe before a court (*habeas corpus*), and have a right to compensation in case of unlawful arrests. Even if the rights of liberty and personal security are often mentioned together in legal sources, they have different scopes. While the right to liberty may be interpreted extensively but remains tied to the physical dimension (Schabas, 2015, p. 226), personal security covers both bodily and mental harm and applies beyond situations of formal deprivation of liberty

[10] UN General Assembly, *Universal Declaration of Human Rights*, 10 December 1948, 217 A (III); UN General Assembly, International Covenant on Civil and Political Rights, 16 December 1966, United Nations, Treaty Series, vol. 999, p. 171. The prohibition of arbitrary detention is also foreseen in Article 5 ECHR, Article 5 CFREU, Article 7 of the American Convention, Article 6 of the African Charter, and Article 14 of the Arab Charter.

(Taylor, 2020, p. 246). For example, the Human Rights Committee has applied the right to security of the person in cases involving death threats, harassment, and various forms of intimidation.[11]

Given this, the right to liberty would arguably not apply to blocking measures in the metaverse for the lack of material apprehension. However, the right to personal security would still be interfered with for the psychological – and potentially physical – impact on users. In particular, blocking measures may have similar legal effects as non-coercive pre-trial measures, like those that prevent suspects from attending specific venues. Therefore, the procedural safeguards imposed by international treaties should apply to metaverse blocking measures.

For their potential physical impact on users, blocking measures should also respect requirements preventing the risk of abuse by the police. The freedom from torture and cruel, inhuman, or degrading treatment is foreseen in several international human rights instruments (Articles 5 UDHR, 7 ICCP, 3 ECHR). Specifically, the United Nations Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (UNCAT) describes torture as any intentional act by a public official that inflicts severe physical or mental suffering, but its Article 1 excludes from this definition any pain arising only from, inherent in or incidental to lawful sanctions. The distinction between ill-treatment and legal measures such as arrests is a matter of concrete evaluation. According to the European Court of Human Rights, police behaviour amounts to torture or other ill-treatment when it attains a minimum level of severity, which is assessed against the circumstances of the case, such as the nature and context of the treatment, the manner and method of its execution, its duration, its physical or mental effects and, in some instances, the sex, age, and state of health of the victim.[12] Extending this reasoning to the metaverse, blocking measures should not be considered illegal per se for their potential consequences on users (i.e., cybersickness). However, States should take precautions to hamper abuses, for instance, limiting their execution to the time strictly necessary to prevent the commission of a serious crime or excluding their application to child users.

In our case, these norms should be integrated into the treaty or the EU Regulation governing the actions of the metaverse police presented in Sect. 4. Such a framework should establish a legal basis for blocking measures, define their scope and safeguards, foresee information rights for blocked users, and identify a competent, independent authority to review blocking measures and decide on compensation if necessary.

Drawing on the analysis developed thus far, the following section provides policy recommendations to begin this process.

[11] *Delgado Páez v Colombia*, Communication No 195/1985 (Application No) UN Doc CCPR/C/39/D/195/1985 [emphasis added]. See also *Gunaratna v Sri Lanka*, no. 1432/2005, UN Doc CCPR/95/D/1432/2005, § 8.4; *Chongwe v Zambia*, no. 821/1998, UN Doc. CCPR/C/79/D/821/1998, § 5.4; *Dias v Angola*, no. 711/1996, UN Doc. CCPR/C/68/D/711/1996, § 8.3; *Jayawardena v Sri Lanka*, no. 916/2000, UN Doc. CCPR/C/675/D/916/2000, § 7.2.

[12] ECtHR, 26 October 2000, *Kudła v. Poland*, App. no. 30,210/96, §§ 91 ff.

Policy Recommendations and Conclusions

In this section, we formulate some policy recommendations (PR) for EU regulators, but also for States beyond the EU, should they take the initiative to regulate policing in the metaverse:

PR_1 *Regulators should recognise the implications of harmful experiences caused by illicit behaviour in the metaverse and adopt specific legislation to criminalise the conducts that generate the most serious harm.* For instance, dedicated resources should be urgently allocated toward empirical research that systematically investigates the extent, nature, and evolving patterns of criminal activity within the metaverse.

PR_2 *The metaverse should be policed through a co-governance mechanism involving private actors and State authorities.* Multiple actors – such as platforms, communities, and LEAs – should act together to police virtual worlds. While providers may be best placed to detect criminal behaviour on their platforms and mitigate harms through code functionalities (like *Horizon Worlds*' 'Safe Zone'), law enforcement will need to supervise their actions, be able to intervene directly, and ensure incidents are followed up on in the analogue world through international cooperation. For instance, the role of communities in policing should be emphasised for addressing minor misconduct, provided that public authorities supervise them. Such a co-governance mechanism should be enacted at the transnational level to ensure that policing efforts are not allocated disproportionally among LEAs or in some areas of the metaverse.

PR_3 *Given the pan-jurisdictional nature of the metaverse, policing should be regulated at the international level (or at least at the EU level for EU Member States).* An international treaty is the best option to ensure equal allocation of responsibilities and avoid excessive policing. Alternatively, as far as the EU is concerned, a Regulation may be the appropriate solution. This legal framework should provide:

- minimum elements of metaverse offenses;
- the composition and tasks of the metaverse police (see also **PR_4**);
- rules to establish jurisdiction *ex-ante* and *ex-post facto* (see also **PR_5**);
- procedural safeguards for police activities (see also **$PR_{6\text{-}7}$**);
- platform obligations.

PR_4 *States should entrust the policing of the metaverse to a unique supranational body (the metaverse police).* This entity should be endowed with autonomous coercive powers and be composed of national officers of the adhering States. Alternatively, national police forces could be directly involved in policing the metaverse, but a supranational body like EUROPOL or INTERPOL should coordinate them.

PR_5 *Regulators should establish specific criteria to solve jurisdictional conflicts in the metaverse.* For crime prevention, digital platforms would fall under the competence of the metaverse police whenever they are established or operate in a State member of the EU, or the international treaty mentioned above (nationality criterion). After the commission of a

criminal offense, the residence or domicile criterion will be the most practical solution to identify the competent body to take action in the analogue world. National LEAs would thus be informed that someone in their territory has committed an offense in the metaverse, allowing them to initiate investigations where appropriate.

PR_6 *Regulators should distinguish private and public places in the metaverse and regulate police surveillance accordingly.* The police should be able to access public metaverse spaces to observe avatars' interactions 'manually', but gathering data in real-time should be subject to stricter conditions. Private metaverse spaces and the data generated therein should be subject to the same protection as the home in physical spaces. They should be accessed by the police only under prior authorisation by an independent authority.

PR_7 *Regulators should protect users' personal security and provide safeguards for metaverse 'arrests' (blocking measures).* Blocking measures cannot be appropriately defined as arrests but have similar fundamental rights implications. While the right to liberty cannot extend to such measures because of the impossibility of material apprehension, their psychological and physical consequences should be covered by the right to personal security and freedom from torture and cruel, inhuman, and degrading treatment.

Before implementing these recommendations, States must decide which illicit behaviours will be criminalised in the immersive environments. Currently, existing laws cannot always extend to crimes committed in the metaverse. Although aggressive virtual behaviours may soon induce physical stimuli in other avatars (Cheong, 2022, p. 16), these sensations would not compare to analogue-world equivalents. Nevertheless, their psychological effects should not be underestimated either. Criminalising behaviours in immersive environments will require creating a metaverse policing system, to which States may choose to adhere. Their decision will be a matter of power and politics. The chance to profit from the metaverse market is a compelling incentive for States (World Economic Forum, 2023a) because social and economic benefits will develop only if users feel safe in this environment. An equal allocation of police responsibilities might be another factor in embracing this system. If States decide not to do so, users will have to realise that some metaverses may be underpoliced or not be policed at all, with very different degrees of safety. The legitimacy of metaverse policing will depend not only on enforcement mechanisms, but on shared norms that treat virtual harm as real.

Acknowledgements We are grateful to Michele Caianiello for his comments on an earlier version of the manuscript.

Author Contributions Although the article is the result of a joint reflection of all the authors, its drafting is broken down as follows: Isadora Neroni Rezende (Sects. 4, 5), Emmie Hine (Sect. 3), Mariarosaria Taddeo (Sect. 1), Luciano Floridi (Sect. 2). All authors have equally contributed to the elaboration and drafting of the recommendations in Sect. 6.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement. No funding was received to assist with the preparation of this manuscript.

Data Availability not applicable.

Declarations

Research Involving Human Participants and/or Animals Not applicable.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethical Approval Not applicable.

Informed Consent Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Adejumo, O. (2022). South Korea sentences metaverse sexual abuser to 4 years imprisonment. *CryptoSlate*. https://cryptoslate.com/south-korea-sentences-metaverse-sexual-abuser-to-4-years-imprisonment/

Agence France-Presse. (2023). New Fortnite mission: Reaching out to abused children. *Technology Inquirer*. https://technology.inquirer.net/101040/new-fortnite-mission-reaching-out-to-abused-children

Ajibola, O. (2022). *The metaverse has a police department for arresting people in VR*. Techbooky.

Altman, I., & Zube, E. H. (Eds.). (1989). *Public places and spaces*. Plenum.

Arrest. (n.d.). In *online etymology dictionary*. https://www.etymonline.com/word/arrest

Ashworth, A. (2006). *Principles of criminal law*. Oxford University Press.

Atzori, M. (2017). Blockchain technology and decentralized governance: Is the state still necessary? *Journal of Governance and Regulation*, *6*(1), 45–62. https://doi.org/10.22495/jgr_v6_i1_p5

Australian Associated Press. (2023). Online gaming platforms such as Roblox used as 'Trojan horse' for extremist recruitment of children, AFP warns. *The Guardian*. https://www.theguardian.com/australia-news/2023/dec/03/online-gaming-platforms-such-as-roblox-used-as-trojan-horse-for-extremist-recruitment-of-children-afp-warns

Babe Howell, K. (2016). The costs of broken windows policing: Twenty years and counting. *Cardozo Law Review*, *37*(3), 1059–1074.

Basu, T. (2021). *The metaverse has a groping problem already*. MIT Technology Review.

Bates, L. (2025). Misogyny in the metaverse: Is Mark Zuckerberg's dream world a no-go area for women? *The Guardian*. https://www.theguardian.com/society/2025/jun/10/the-misogyny-of-the-metaverse-is-mark-zuckerbergs-dream-world-a-no-go-area-for-women

BBC. (2007). *Virtual theft leads to arrest*. http://news.bbc.co.uk/2/hi/7094764.stm

Behr, K. M., Nosper, A., Klimmt, C., & Hartmann, T. (2005). Some practical considerations of ethical issues in VR research. *Presence: Teleoperators and Virtual Environments,14*(6), 668–676. https://doi.org/10.1162/105474605775196535

Beres, N. A., Frommel, J., Reid, E., Mandryk, R. L., & Klarkowski, M. (2021). Don't you know that you're toxic: Normalization of toxicity in online gaming. Proceedings of the 2021 CHI conference on human factors in computing systems, 1–15. https://doi.org/10.1145/3411764.3445157

Brenner, S. (2001). Is there such a thing as virtual crime?? *California Criminal Law Review*, *4*, 1–72.

Brenner, S. W. (2008). Fantasy Crime: The Role of Criminal Law in Virtual Worlds. *Vanderbilt Journal of Entertainment & Technology Law, 11*(1)

Cadet, L. B., & Chainay, H. (2020). Memory of virtual experiences: Role of immersion, emotion and sense of presence. *International Journal of Human-Computer Studies,144*, Article 102506. https://doi.org/10.1016/j.ijhcs.2020.102506

Casey, P., Baggili, I., & Yarramreddy, A. (2021). Immersive virtual reality attacks and the human joystick. *IEEE Transactions on Dependable and Secure Computing,18*(2), 550–562. https://doi.org/10.1109/TDSC.2019.2907942

Centre for Countering Digital Hate. (2023). *Horizon worlds exposed*. https://counterhate.com/research/horizon-worlds-exposed/

Centre for Countering Digital Hate. (2021). *New research shows Metaverse is not safe for kids*. https://counterhate.com/blog/new-research-shows-metaverse-is-not-safe-for-kids/

Cheong, B. C. (2022). Avatars in the metaverse: Potential legal issues and remedies. *International Cybersecurity Law Review,3*(2), 467–494. https://doi.org/10.1365/s43439-022-00056-9

Cieslak, M., & Gerken, T. (2023). *Interpol working out how to police the metaverse*. https://www.bbc.com/news/technology-64501726

Cohen, J. (2013). What privacy is for. *Harvard Law Review,126*(7), 1904–1933.

Collister, L. B. (2014). Surveillance and community: Language policing and empowerment in a world of Warcraft guild. *Surveillance & Society,12*(3), 337–348. https://doi.org/10.24908/ss.v12i3.4956

Corse, A. (2023). *Twitter missed dozens of known images of child sexual abuse material, researchers say*. https://www.wsj.com/articles/twitter-missed-dozens-of-known-images-of-child-sexual-abuse-material-researchers-say-58d44f7b

Cox, K. E. (1999). Beyond self-defense: United Nations peacekeeping operations & (and) the use of force. *Denver Journal of International Law and Policy,27*(2), 239–273.

Crawford, A., & Smith, T. (2022). *Metaverse app allows kids into virtual strip clubs*. BBC. https://www.bbc.com/news/technology-60415317?uuid=qFhIDbLg6IVsy6Lp2152

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough?? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology,19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

De Almeida, G. G. F. (2023). Cities and territorial brand in the metaverse: The metaverse SEOUL case. *Sustainability,15*(13), Article 10116. https://doi.org/10.3390/su151310116

De Back, T. T., Tinga, A. M., & Louwerse, M. M. (2024). Natural- and redirected walking in virtual reality: Spatial performance and user experience. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-19879-1

Deflem, M. (2000). Bureaucratization and social control: Historical foundations of international police cooperation. *Law & Society Review,34*(3), 739. https://doi.org/10.2307/3115142

Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M. M., Dennehy, D., Metri, B., Buhalis, D., Cheung, C. M. K., Conboy, K., Doyle, R., Dubey, R., Dutot, V., Felix, R., Goyal, D. P., Gustafsson, A., Hinsch, C., Jebabli, I., & Wamba, S. F. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management,66*, Article 102542. https://doi.org/10.1016/j.ijinfomgt.2022.102542

Easterbrook, F. H. (1996). *Cyberspace and the law of the horse* (pp. 207–216). The university of chicago legal forum.

Ehrett, J. S. (2016). E-judiciaries: A model for community policing in cyberspace. *Information & Communications Technology Law,25*(3), 272–291. https://doi.org/10.1080/13600834.2016.1236428

European Commission. (2020). Impact assessment accompanying the document proposal for a regulation of the european parliament and of the council on a single market for digital services (Digital Services Act) and amending directive 2000/31/EC. https://doi.org/10.5040/9781782258674

European Commission. (2022). Impact assessment report accompanying the document proposal for a regulation of the European parliament and the council laying down rules to prevent and combat child sexual abuse.

European Commission. (2023). Virtual worlds fit for people. https://digital-strategy.ec.europa.eu/en/policies/virtual-worlds

European Roma Rights Centre. (2023). New research points to institutional racism against Roma in North Macedonia's criminal justice system. www.errc.org/press-releases/new-research-points-to-institutional-racism-against-roma-in-north-macedonias-criminal-justice-system

EUROPOL. (2022). *Policing in the metaverse: What law enforcement needs to know : An observatory report from the Europol innovation lab*. Publications Office. https://data.europa.eu/doi/ https://doi.org/10.2813/81062

Falchuk, B., Loeb, S., & Neff, R. (2018). The social metaverse: Battle for privacy. *IEEE Technology and Society Magazine*, *37*(2), 52–61. https://doi.org/10.1109/MTS.2018.2826060

Farrant, T. (2024). *UK police launch first investigation into virtual rape in metaverse*. Euronews. https://www.euronews.com/next/2024/01/04/british-police-launch-first-investigation-into-virtual-rape-in-metaverse

Floridi, L. (2008). The method of levels of abstraction. *Minds and Machines*, *18*(3), 303–329. https://doi.org/10.1007/s11023-008-9113-7

Floridi, L. (2020). The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & Technology,33*(3), 369–378. https://doi.org/10.1007/s13347-020-00423-6

Floridi, L. (2021). The European legislation on AI: A brief analysis of its philosophical approach. *Philosophy & Technology,34*(2), 215–222. https://doi.org/10.1007/s13347-021-00460-9

Floridi, L. (2022). Metaverse: A matter of experience. *Philosophy & Technology*, *35*(3), 73, s13347-022-00568–6. https://doi.org/10.1007/s13347-022-00568-6

Fortis, S. (2023). *China declares stealing digital collections like NFTs liable for criminal theft sentence*. Cointelegraph. https://cointelegraph.com/news/china-declares-stealing-digital-collections-like-nfts-liable-for-criminal-theft-sentence

France-Presse, A. (2022). *Russian teenager jailed over 'Minecraft plot to blow up virtual spy HQ*.' The Guardian. https://www.theguardian.com/world/2022/feb/10/russian-teenager-nikita-uvarov-jailed-over-minecraft-plot-to-blow-up-virtual-spy-hq

Garner, B. A. (2014). Arrest. In *Black's law dictionary* (10th edition).

Garon, J. M. (2022). Legal implications of a ubiquitous metaverse and a Web3 future. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4002551

Gómez-Quintero, J., Johnson, S. D., Borrion, H., & Lundrigan, S. (2024). A scoping study of crime facilitated by the metaverse. *Futures,157*, Article 103338. https://doi.org/10.1016/j.futures.2024.103338

Goujard, C. (2024). *Meta, TikTok take EU to court over online content rulebook*. POLITICO. https://www.politico.eu/article/tiktok-joins-meta-in-suing-eu-over-online-content-rulebook/

Gray, J. E., Carter, M., & Egliston, B. (2024). Trust and safety in social VR: Current industry practices. In J. E. Gray, M. Carter, & B. Egliston (Eds.), *Governing social virtual reality: preparing for the content, conduct and design challenges of immersive social media* (pp. 61–75). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-61831-4_6

Greenberg, A. (2022). *Inside the bitcoin bust that took down the web's biggest child abuse site*. https://www.wired.com/story/tracers-in-the-dark-welcome-to-video-crypto-anonymity-myth/

Gromer, D., Reinke, M., Christner, I., & Pauli, P. (2019). Causal interactive links between presence and fear in virtual reality height exposure. *Frontiers in Psychology,10*, Article 141. https://doi.org/10.3389/fpsyg.2019.00141

Haber, E. (2024). The criminal metaverse. *Indiana Law Journal*, *99*(3). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4400281

Habermas, J. (1991). *The structural transformation of the public sphere* (1st ed.). MIT Press.

Headleand, C. J., Free, J., Farndale, S., & Hall, M. (2020). Virtual community support officers: Community policing in the digital space. 2020 international conference on cyberworlds (CW), 121–124. https://doi.org/10.1109/CW49994.2020.00025

Hillyard, P., & Tombs, S. (2007). From 'crime' to social harm? *Crime, Law and Social Change,48*(1–2), 9–25. https://doi.org/10.1007/s10611-007-9079-z

Hillyard, P., & Tombs, S. (2021). Beyond criminology? In P. Davies, P. Leighton, & T. Wyatt (Eds.), *The palgrave handbook of social harm* (pp. 11–36). Springer International Publishing. https://doi.org/10.1007/978-3-030-72408-5_2

Hine, E. (2023). *Content moderation in the metaverse could be a new frontier to attack freedom of expression*. *Philosophy & Technology*. https://doi.org/10.1007/s13347-023-00645-4

Hine, E., Rezende, I. N., Roberts, H., Wong, D., Taddeo, M., & Floridi, L. (2024). Safety and privacy in immersive extended reality: An analysis and policy recommendations. *Digital Society*, *3*(2), 33. https://doi.org/10.1007/s44206-024-00114-1

Horwitz, J., & Blunt, K. (2023). Instagram connects vast pedophile network. *The Wall Street Journal*. https://www.wsj.com/articles/instagram-vast-pedophile-network-4ab7189

INTERPOL. (2024). *Metaverse: A law enforcerment perspective*.

Jackson, J., McKay, T., Cheliotis, L., Bradford, B., Fine, A., & Trinkner, R. (2023). Centering race in procedural justice theory: Structural racism and the under- and overpolicing of black communities. *Law and Human Behavior,47*(1), 68–82.

Keiler, J., & Roef, D. (2019). Principles of criminalization and the limits of criminal law. In *comparative concepts of criminal law* (3rd ed., pp. 35–83). Intersentia.

Knight, W. (2005). *Computer characters mugged in virtual crime spree*. New Scientist. https://web.archive.org/web/20231108210055/https://www.newscientist.com/article/dn7865-computer-characters-mugged-in-virtual-crime-spree/

Laue, C. (2011). Crime potential of metaverses. In K. Cornelius & D. Hermann (Eds.), *Virtual worlds and criminality* (pp. 19–29). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20823-2_2

Leenes, R. (2008). Privacy in the metaverse. In S. Fischer-Hübner, P. Duquenoy, A. Zuccato, & L. Martucci (Eds.), *The future of identity in the information society* (pp. 95–112). Springer US. https://doi.org/10.1007/978-0-387-79026-8_7

Lemley, M. A., & Volokh, E. (2017). Law, virtual reality, and augmented reality. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2933867

Lessig, L. (1999). The law of the horse: What cyberlaw might teach. *Harvard Law Review,113*(2), 501. https://doi.org/10.2307/1342331

Mackenzie, S. (2022). Criminology towards the metaverse: Cryptocurrency scams, grey economy and the technosocial. *The British Journal of Criminology*, *62*(6), 1537–1552. https://doi.org/10.1093/bjc/azab118

Madiega, T., Car, P., Niestadt, M., & Van de Pol, L. (2022). *Metaverse: Opportunities, risks and policy implications*. *European Parliamentary Research Service*. https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733557/EPRS_BRI(2022)733557_EN.pdf

Makransky, G., & Mayer, R. E. (2022). Benefits of taking a virtual field trip in immersive virtual reality: Evidence for the immersion principle in multimedia learning. *Educational Psychology Review,34*(3), 1771–1798. https://doi.org/10.1007/s10648-022-09675-4

Maresca, T. (2023). *Seoul launches ambitious metaverse platform for city services, tourism*. UPI. https://www.upi.com/Top_News/World-News/2023/01/16/Seoul-Metaverse-virtual-platform-launched/6531673859297/

Martens, M. A., Antley, A., Freeman, D., Slater, M., Harrison, P. J., & Tunbridge, E. M. (2019). It feels real: Physiological responses to a stressful virtual reality environment and its impact on working memory. *Journal of Psychopharmacology,33*(10), 1264–1273. https://doi.org/10.1177/0269881119860156

Marthews, A., & Tucker, C. (2017). The impact of online surveillance on behavior. In D. Gray & S. E. Henderson (Eds.), *The cambridge handbook of surveillance law* (1st ed., pp. 437–454). Cambridge University Press. https://doi.org/10.1017/9781316481127.019

Meta Horizon Worlds Frequently Asked Questions | Quest help | Meta Store. (n.d.). Retrieved January 28, 2025, from https://www.meta.com/en-gb/help/quest/articles/horizon/explore-horizon-worlds/horizon-frequently-asked-questions/

Meta. (2024). *Moderators and community guides in meta horizon worlds*. https://www.meta.com/help/quest/articles/horizon/safety-and-privacy-in-horizon-worlds/community-guides-in-horizon/

Michalowski, R. J. (2016). What is crime?? *Critical Criminology,24*(2), 181–199. https://doi.org/10.1007/s10612-015-9303-6

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems*, *E77-D*(12).

Murray, C. D., & Sixsmith, J. (1999). The corporeal body in virtual reality. *Ethos,27*(3), 315–343.

Mütterlein, J. (2018). The three pillars of virtual reality? Investigating the roles of immersion, presence, and interactivity. Proceedings of the 51st Hawaii international conference on system sciences, 1407–1415. http://hdl.handle.net/10125/50061

Nabben, K. (2021). Building the metaverse: 'Crypto states' and corporates compete, down to the hardware. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3981345

Nichols, S. (2022). *Metaverse rollout brings new security risks, challenges | TechTarget*. TechTarget. https://www.techtarget.com/searchsecurity/news/252513072/Metaverse-rollout-brings-new-security-risks-challenges

Oppenheim, M. (2022). *Woman reveals 'nightmare' of being 'gang raped' in virtual reality*. The Independent. https://www.independent.co.uk/news/uk/home-news/metaverse-gang-rape-virtual-world-b2005959.html

Outlaw, J. (2018). *Virtual harassment: The social experience of 600+regular virtual reality (VR) users*. VirtualRealityPop. https://virtualrealitypop.com/virtual-harassment-the-social-experience-of-600-regular-virtual-reality-vr-users-23b1b4ef884e

Oxman, B. H. (2007). Jurisdiction of States. *Max planck encyclopedia of public international law*. Oxford University Press.

Pemberton, S. (2007). Social harm future(s): Exploring the potential of the social harm approach. *Crime, Law and Social Change,48*(1–2), 27–41. https://doi.org/10.1007/s10611-007-9078-0

Qin, H. X., Wang, Y., & Hui, P. (2025). Identity, crimes, and law enforcement in the metaverse. *Humanities and Social Sciences Communications,12*(1), Article 194. https://doi.org/10.1057/s41599-024-04266-w

Qiwei, L., Zhang, S., Kasper, A. T., Ashkinaze, J., Eaton, A. A., Schoenebeck, S., & Gilbert, E. (2024). Reporting non-consensual intimate media: An audit study of deepfakes (arXiv:2409.12138). arXiv. https://doi.org/10.48550/arXiv.2409.12138

Regan, P. M. (2015). Privacy and the common good: Revisited. In B. Roessler & D. Mokrosinska (Eds.), *Social dimensions of privacy* (1st ed., pp. 50–70). Cambridge University Press. https://doi.org/10.1017/CBO9781107280557.004

Riis, T., & Schwemer, S. F. (2018). Leaving the European safe harbor, sailing towards algorithmic content regulation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3300159

Robinson, C. D., & Scaglion, R. (1987). The origin and evolution of the police function in society: Notes toward a theory. *Law & Society Review,21*(1), 109. https://doi.org/10.2307/3053387

Ruppert, E. S. (2006). Rights to public space: Regulatory reconfigurations of liberty. *Urban Geography,27*(3), 271–292. https://doi.org/10.2747/0272-3638.27.3.271

Sales, N. J. (2024). *A girl was allegedly raped in the metaverse. Is this the beginning of a dark new future*? The Guardian.

Santana, C., & Albareda, L. (2022). Blockchain and the emergence of decentralized autonomous organizations (DAOs): An integrative model and research agenda. *Technological Forecasting and Social Change,182*, Article 121806. https://doi.org/10.1016/j.techfore.2022.121806

Schabas, W. (2015). *European convention on human rights: A commentary* (1st ed.). Oxford University Press.

Schmitt, M. N. (Ed.). (2017). Jurisdiction. *Tallinn manual 2.0 on the international law applicable to cyber operations* (2nd ed., pp. 51–78). Cambridge University Press. https://doi.org/10.1017/9781316822524.009

Schulenberg, K., Li, L., Freeman, G., Zamanifard, S., & McNeese, N. J. (2023). Towards leveraging AI-based moderation to address emergent harassment in social virtual reality. Proceedings of the 2023 CHI conference on human factors in computing systems, 1–17. https://doi.org/10.1145/3544548.3581090

Sethi, A. (2022). Security and privacy in metaverse: Issues, challenges, and future opportunities. 2.

Singapore Ministry of Home Affairs. (2023). Issuance of orders under the internal security act against two self-radicalised Singaporean youths. Ministry of home affairs. http://www.mha.gov.sg/mediaroom/press-releases/issuance-of-orders-under-the-internal-security-act-against-two-self-radicalised-singaporean-youths/

Stevenson, A. (Ed.). (2010). *Oxford dictionary of english* (3rd ed.). Oxford University Press.

SumOfUs (2022). *Metaverse: Another cesspool of toxic content*. https://www.eko.org/images/Metaverse_report_May_2022.pdf

Taylor, P. M. (2020). *A commentary on the international covenant on civil and political rights: The UN human rights committee's monitoring of ICCPR rights* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108689458

Toolkit: Cross-Border Content Moderation. (2021). Internet & jurisdiction policy network. www.internetjurisdiction.net/content/toolkit

Tseng, W. J., Bonnail, E., McGill, M., Khamis, M., Lecolinet, E., Huron, S., & Gugenheimer, J. (2022). The dark side of perceptual manipulations in virtual reality. *CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3491102.3517728

Urban, J. M., Karaganis, J., & Schofield, B. L. (2017). *Notice and takedown in everyday practice*. Berkeley Law.

Wang, H. (2023). How to deal with virtual property crime: Judicial dilemma and a theoretical solution from China. *Computer Law & Security Review,49*, Article 105808. https://doi.org/10.1016/j.clsr.2023.105808

Wang, Y., Su, Z., Zhang, N., Xing, R., Liu, D., Luan, T. H., & Shen, X. (2022). A survey on metaverse: Fundamentals, security, and privacy. *IEEE Communications Surveys & Tutorials*, 1–1. https://doi.org/10.1109/COMST.2022.3202047

World Economic Forum (2023a). *Metaverse: What are the economic benefits? | World Economic Forum*. https://www.weforum.org/agenda/2023/06/what-will-be-the-economic-benefits-of-the-metaverse/

World Economic Forum, J. U. (2023b). Metaverse Privacy and Safety.

Zallio, M., & Clarkson, P. J. (2022). Designing the metaverse: A study on inclusion, diversity, equity, accessibility and safety for digital immersive environments. *Telematics and Informatics,75*, Article 101909. https://doi.org/10.1016/j.tele.2022.101909

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.