AI and Ethics
https://doi.org/10.1007/s43681-024-00450-5

ORIGINAL RESEARCH

Ethical implications of AI in the Metaverse

Alesia Zhuk[1]

Received: 8 February 2024 / Accepted: 21 February 2024

Abstract
This paper delves into the ethical implications of AI in the Metaverse through the analysis of real-world case studies, including Horizon Worlds, Decentraland, Roblox, Sansar, and Rec Room. The examination reveals recurring concerns related to content moderation, emphasising the need for a human-AI hybrid approach to strike a balance between creative freedom and user safety. Privacy and data protection emerge as crucial considerations, highlighting the importance of transparent communication and user data control for responsible AI implementation. Additionally, promoting inclusivity and diversity is emphasised, calling for transparent governance, diverse representation, and collaboration with ethics experts to ensure equitable AI practices. By addressing these specific ethical challenges, we can pave the way towards a responsible and user-centric Metaverse, maximising its potential while safeguarding user well-being and rights.

Keywords Virtual environments · Content curation · User privacy · Governance mechanisms · User-centric design · Creative freedom

1 Introduction

The Metaverse, a virtual world encompassing immersive experiences and social interactions, is rapidly evolving with the integration of artificial intelligence (AI) technologies. As AI becomes increasingly prevalent in shaping the virtual landscape, it raises significant ethical concerns that demand attention and deliberation. This paper explores the ethical implications of AI in the Metaverse, focusing on issues of bias, discrimination, privacy, and transparency.

According to Moneta [60], the Metaverse refers to a collective virtual shared space that goes beyond traditional virtual reality experiences. It encompasses a vast interconnected network of virtual worlds, augmented reality overlays, and digital platforms [92], where users engage with each other and their surroundings in real-time. The immersive and interactive nature of the Metaverse offers unprecedented opportunities for socialisation, commerce, and self-expression [88].

While AI offers boundless potential for shaping the Metaverse, it also necessitates critical reflection on ethical implications. The utilisation of AI-driven personalisation, data collection, and algorithmic decision-making raises concerns about privacy, bias, and individual autonomy (Nasar and Kamal [63]. It is imperative to ensure that AI technologies within the Metaverse adhere to ethical principles, fostering an inclusive, safe, and equitable virtual space [8].

To address these ethical challenges, a legal and regulatory framework is essential. However, the existing regulatory landscape for AI in virtual environments remains nascent [80]. Researchers such as Bang and Kim [6], Benjamins et al. [8] and Habbal et al. [39] argue that the unique characteristics of the Metaverse necessitate tailored regulations that address the ethical implications of AI technologies. Robust legal and ethical guidelines are required to foster responsible AI practices within the evolving virtual world [26, 38].

In light of these concerns, this paper aims to contribute to the ongoing discourse surrounding the ethical implications of AI in the Metaverse. As the Metaverse continues to evolve, the integration of AI technologies demands a balanced approach that considers the potential benefits alongside the ethical risks. By taking proactive steps to address bias, discrimination, privacy concerns, and transparency issues, it is possible to cultivate an ethically sound and sustainable virtual world. The development of a comprehensive regulatory framework and adherence to ethical guidelines

✉ Alesia Zhuk
alesia.zhuk@ug.uchile.cl

[1] Law Department, Pompeu Fabra University, Barcelona, Spain

Published online: 13 March 2024

are essential to foster a Metaverse that prioritises user welfare and supports responsible AI practices.

2 Ethical challenges of AI in the metaverse

As AI technologies become increasingly integrated within the Metaverse, the virtual world presents a unique set of ethical challenges. This chapter explores the ethical implications of AI in the Metaverse, focusing on the complex issues of bias, discrimination, privacy, and transparency. By examining these challenges, we aim to shed light on the intricate ethical landscape that emerges as AI plays a prominent role in shaping virtual environments.

The integration of AI technologies within the Metaverse introduces complex ethical challenges. One significant concern is the potential for bias and discrimination in AI algorithms. Studies conducted by Buolamwini and Gebru [12], Ferrer et al. [35] and Varona & Suárez [89] have demonstrated that AI systems can inherit biases present in the data they are trained on, leading to discriminatory outcomes. In the Metaverse, biased AI algorithms can perpetuate inequality and marginalisation, affecting user experiences and interactions [3].

Moreover, privacy considerations arise as AI systems within the Metaverse collect vast amounts of user data [43, 68]. The ability to monitor and analyse user behavior raises concerns about surveillance and data protection [25, 26, 57]. With the potential for data breaches and unauthorised access, users may face privacy risks, as highlighted by O'Brolcháin et al. [65] and Huang et al. [42] in their studies on virtual reality, social networks and privacy implications.

Transparency and explainability are crucial factors for building trust in AI systems [5]. However, AI algorithms employed in the Metaverse often operate as black boxes, making it difficult for users to understand the decision-making process behind AI-driven interactions [39, 43]. Research by Mittelstadt et al. [59] emphasises the need for explainable AI models to ensure transparency, accountability, and user comprehension in the virtual world.

In this chapter, a deep dive will be taken into these ethical challenges associated with AI in the Metaverse, exploring their implications and considering potential solutions. Through a thoughtful examination of bias, discrimination, privacy, and transparency, progress can be made towards a more ethical and inclusive Metaverse experience for all users.

2.1 Bias and discrimination

AI algorithms in the Metaverse have the potential to perpetuate biases and contribute to discriminatory outcomes, raising significant ethical concerns. As AI systems learn from vast amounts of data, biases present in that data can be inadvertently replicated and reinforced, leading to biased decision-making processes within virtual environments [18, 90].

A groundbreaking study by Buolamwini and Gebru [12] on facial recognition algorithms revealed substantial racial and gender biases. They found that commercial gender classification systems exhibited higher error rates for darker-skinned and female faces. These biases can have severe consequences within the Metaverse, where user avatars and virtual representations rely on facial recognition technology. Biased algorithms can perpetuate inequality and marginalisation, adversely affecting user experiences and social interactions.

In the context of content curation and recommendation algorithms, bias and discrimination can also emerge. Sap et al. [78] conducted research on major platforms and found that algorithmic systems for content recommendation inadvertently reinforced stereotypes and biases. This can limit the diversity of information and perspectives accessible to users in the Metaverse, contributing to echo chambers and further entrenching biases.

Addressing bias and discrimination in AI algorithms within the Metaverse requires careful consideration and proactive measures. One approach is to ensure diverse and representative training datasets [75]. Mittelstadt et al. [59] highlight the importance of comprehensive training data that reflects the diversity of user populations, avoiding underrepresentation or marginalisation of certain groups. By incorporating diverse perspectives during the data collection process, developers can mitigate biases that may arise [75].

Ongoing monitoring and evaluation of AI algorithms' performance are essential to detect and rectify biases [36, 75]. Regular audits and transparency in algorithmic decision-making processes, as advocated by Selbst et al. [83] and Lee et al. [49], can help identify and correct biases, promoting fairness and equity within the virtual world. Transparent reporting of AI development and deployment processes can foster trust among users, enabling them to hold AI systems accountable [34],Brundage et al. [11].

Additionally, interdisciplinary collaboration is crucial to address bias and discrimination in AI algorithms [35]. It requires efforts from AI developers, platform operators, policymakers, and ethicists to collectively work towards fair and unbiased AI systems. Guidelines and regulations should be established to ensure responsible AI development and usage, incorporating principles of fairness, accountability, and transparency [34, 79].

2.2 Privacy and data security concerns

The integration of AI algorithms in the Metaverse raises significant privacy concerns, encompassing data collection and surveillance practices within virtual environments [39].

Data collection practices within virtual environments have raised concerns regarding user privacy and data security [1]. Research by Mystakidis [61] explores data collection practices within the Metaverse and emphasises the need for transparent and accountable approaches to data handling. It highlights the potential privacy risks associated with extensive data collection and the need for user awareness and control over their personal information.

The use of AI in the Metaverse can introduce privacy risks and potential violations. Studies by Huang et al. [42] discuss the privacy challenges posed by immersive technologies, including the Metaverse, such as the risks of unauthorised surveillance, data breaches, and unintended data disclosures. They emphasise the importance of robust privacy frameworks and proactive measures to protect user privacy in virtual environments.

User consent mechanisms and data protection practices play a crucial role in addressing privacy concerns in the Metaverse. Bavana [7] evaluates the challenges of obtaining informed consent from users within virtual environments and proposes privacy-enhancing solutions. They emphasise the importance of user control over personal data and the incorporation of privacy-by-design principles to ensure data protection in the Metaverse.

Moreover, studies by Bozkir et al. [10] and Lammerding et al. [46] explore the privacy implications of AI-driven technologies, including virtual reality (VR) and augmented reality (AR), which are integral components of the Metaverse. These studies underline the significance of user awareness, transparent data practices, and user-friendly privacy controls to safeguard personal data and uphold user privacy within immersive virtual environments.

The findings from these studies highlight the need for robust privacy measures in the Metaverse. Transparent and accountable data collection practices, privacy-enhancing solutions, and user-centric privacy controls are essential to protect user privacy within virtual environments.

2.3 Transparency and explainability

Transparency and explainability are critical pillars in the successful integration of AI systems within the Metaverse [39]. As AI plays an increasingly prominent role in shaping virtual experiences, the need for transparent AI systems becomes paramount. Research conducted by Veale et al. [91] emphasises that transparent AI systems enable users to understand the rationale behind AI-driven decisions. This level of transparency fosters trust, as users can comprehend how choices are made, empowering them to make informed decisions and mitigating potential biases and discriminatory outcomes.

In the context of complex virtual environments, understanding and interpreting AI decision-making processes pose significant challenges (Luck & Ayllet [54]. As highlighted by Mittelstadt et al. [59], achieving interpretability in AI algorithms within the Metaverse is crucial. Bridging the gap between AI developers and users becomes essential to enable users to grasp the underlying logic behind AI-driven decisions. When users can access explanations for AI outcomes, they can engage meaningfully with the virtual world and have a clearer understanding of how these decisions impact their experiences.

To enhance transparency and explainability in the virtual world, researchers have proposed various approaches. Burrell [14] suggests the development of "transparent by design" AI systems. These systems are designed with built-in mechanisms that facilitate user understanding and influence over AI decision-making processes from the outset. By incorporating transparency as a core principle, these AI systems empower users to be active participants in the virtual world's AI-driven processes.

Furthermore, advances in interpretable AI techniques have shown promise in providing insights into AI processes. Techniques such as rule-based systems and visual explanations, as explored by Ribeiro et al. [73], aim to shed light on the "black box" nature of AI algorithms. Rule-based systems present decision-making rules in a human-readable format, while visual explanations offer intuitive visualisations of AI's decision pathways. These approaches enhance user comprehension, making AI decision-making more accessible and understandable.

Moreover, transparency and explainability also play a crucial role in addressing accountability and responsibility in the Metaverse [39]. In instances where AI decisions impact users' well-being or rights, the ability to understand and scrutinise those decisions becomes essential. Transparent AI systems facilitate the identification of errors or biases [34], allowing for timely corrections and accountability for any unintended consequences.

Overall, transparent AI systems foster trust, enable user agency, and prevent potential ethical issues [34]. The development of "transparent by design" approaches and interpretable AI techniques ensures that users can engage confidently with AI systems and understand the decisions that influence their virtual experiences.

3 Legal and regulatory considerations

Legal and regulatory frameworks play a crucial role in addressing the ethical challenges associated with AI in virtual environments [16, 66]. As AI technologies continue to advance and become increasingly integrated into virtual environments, it is essential to evaluate the existing legal frameworks to ensure they are equipped to address the ethical implications that arise. This section aims to examine

the current legal frameworks, assess their adequacy in addressing ethical concerns, and explore potential regulatory approaches and policy recommendations for promoting responsible AI use in virtual environments.

3.1 Review of existing legal frameworks for AI in virtual environments

Existing legal frameworks provide a foundation for regulating AI in virtual environments, albeit with varying levels of specificity [16]. These frameworks reflect the recognition of the need to address the ethical implications of AI technologies within virtual environments and strive to ensure the protection of user rights and interests [66]. This section presents a comprehensive review of notable examples of legal frameworks that have implications for AI integration in virtual environments. The analysis aims to shed light on the extent to which these frameworks accommodate the unique ethical challenges posed by AI within the dynamic and immersive virtual world.

The General Data Protection Regulation (GDPR) in the European Union (EU) is a comprehensive regulation that sets standards for data protection and privacy. It applies to AI systems operating within virtual environments that involve the processing of personal data (European [30]. The GDPR emphasises principles such as consent, transparency, and accountability, aiming to safeguard individuals' privacy rights and regulate the collection, storage, and use of personal data within virtual environments (Regulation (EU) 2016/679, [72].

In China, the Data Security Law of the People's Republic of China addresses data protection and cybersecurity concerns, encompassing AI-related activities within virtual environments. This law establishes requirements for data protection, data localisation, and the prevention of cybersecurity threats, ensuring the secure and responsible use of AI technologies within virtual environments [64].

In the United States, the legal landscape for AI in virtual environments is shaped by various laws and regulations, including the Federal Trade Commission (FTC) Act. The FTC Act empowers the Federal Trade Commission to regulate unfair or deceptive practices, including those related to AI systems operating in virtual environments [33]. Additionally, sector-specific laws, such as the Health Insurance Portability and Accountability Act (HIPAA) and the Children's Online Privacy Protection Act (COPPA), provide guidelines for privacy and data protection in specific contexts [44].

While these existing legal frameworks provide a basis for regulating AI in virtual environments, it is important to continually evaluate their adequacy in addressing the evolving ethical concerns. This ensures that the legal landscape remains responsive and effective in addressing the ethical implications of AI use within virtual environments.

3.2 Analysis of the adequacy of current regulations in addressing ethical concerns

The existing legal frameworks for AI in virtual environments lay the foundation for regulation, but their effectiveness in addressing evolving ethical concerns requires careful evaluation (O'Sullivan [66].

The GDPR establishes high standards for data protection and privacy, covering AI systems operating in virtual environments. It emphasises user consent, data transparency, and accountability, providing a robust framework for safeguarding individuals' privacy rights (Regulation (EU) 2016/679, [72]. However, with the continuous advancement of AI technologies, new data processing methods and data-sharing practices may emerge, necessitating regular updates and adaptation of the regulation to maintain its efficacy in protecting user privacy within the Metaverse [69].

Similarly, the Data Security Law of the People's Republic of China mandates data localisation and aims to prevent cybersecurity threats, safeguarding user data and maintaining the integrity of virtual interactions [64]. However, given the ever-evolving nature of cybersecurity threats, continuous monitoring and updates to the regulation may be necessary to stay ahead of potential risks in the dynamic virtual landscape.

Although the FTC Act in the United States offers a mechanism to address harmful AI-driven interactions, its broad language may need further refinement to explicitly encompass AI-specific ethical concerns unique to virtual environments (O'Sullivan, [66]. Clear guidelines for AI system transparency and explainability could enhance the act's effectiveness in preventing misleading or harmful AI interactions.

Moreover, sector-specific laws, such as the HIPAA and the COPPA provide guidelines for privacy and data protection in specific contexts, particularly within healthcare and children's online privacy domains. However, as AI continues to integrate into the Metaverse beyond these specific sectors, more comprehensive regulations are needed to address AI ethics holistically within virtual environments.

Despite these existing legal frameworks providing some level of protection, concerns persist regarding their adequacy in addressing the ethical implications of AI in virtual environments. Notably, biases and discrimination present challenges, as demonstrated by Buolamwini and Gebru's [12] research on facial recognition algorithms, which revealed biases disproportionately affecting certain demographic groups. These biases have the potential to perpetuate inequality and discrimination within virtual environments. However, current regulations may not explicitly address the specific challenges posed by biased AI systems within virtual environments.

Furthermore, the rapidly evolving nature of AI technology and virtual environments poses challenges for existing regulations to keep pace with emerging ethical concerns. As virtual environments advance, new ethical challenges may arise, demanding regulatory responses (Maloney [56]. The dynamic and complex nature of AI in virtual environments calls for ongoing evaluation and updates to ensure that regulations adequately address the ethical implications associated with AI use.

To address these concerns, regulatory bodies must conduct a comprehensive analysis of the current legal frameworks. This analysis should identify areas where regulations may fall short in addressing ethical concerns, such as bias, discrimination, privacy, transparency, and accountability. By identifying these gaps, a more targeted and effective regulations that address the unique ethical challenges of AI in virtual environments can be developed and implemented [95].

3.3 Discussion of potential regulatory approaches and policy recommendations for responsible AI use

To promote responsible AI use in virtual environments, potential regulatory approaches and policy recommendations can be considered. These approaches aim to address the ethical concerns associated with AI technologies and ensure the responsible deployment and use of AI in virtual environments.

One approach is to develop specific regulations or guidelines tailored to AI in virtual environments [31]. These regulations should take into account the unique ethical challenges posed by virtual environments and AI systems operating within them. The establishment of standards for transparency and explainability in AI systems is crucial [5]. Users should have a clear understanding of how AI algorithms operate and make decisions [48]. The European Commission's High-Level Expert Group on AI [29] has proposed ethical guidelines that emphasise transparency, accountability, and user empowerment, providing a framework for responsible AI development and deployment.

Enhancing accountability and oversight of AI systems in virtual environments is another crucial aspect [27]. This can be achieved through the establishment of auditing mechanisms and impact assessments [32]. Audits can evaluate the compliance of AI systems with ethical guidelines and regulatory requirements. Impact assessments help identify potential risks and ethical implications associated with AI use in virtual environments. Multidisciplinary perspectives, involving policymakers, industry experts, and ethicists, should be integrated to inform the development of robust regulatory frameworks that consider the societal impact of AI in virtual environments (Brundage et al. [11].

International cooperation and collaboration are essential in addressing the global nature of AI in virtual environments [19]. Harmonising regulatory approaches and sharing best practices can facilitate the responsible and ethical deployment of AI systems across different jurisdictions (de Almeida [23]. Initiatives like the Global Partnership on Artificial Intelligence (GPAI) bring together countries and organisations to collaborate on AI governance and promote the development of responsible AI practices [37].

By implementing these regulatory approaches and policy recommendations, policymakers and stakeholders can foster responsible AI use in virtual environments. These frameworks aim to ensure transparency, accountability, and user empowerment while addressing the ethical concerns associated with AI deployment. Additionally, they facilitate international cooperation, promoting a consistent and ethical approach to AI governance in virtual environments.

4 Ethical guidelines for AI in the metaverse

The development and adoption of ethical guidelines for the AI use in Metaverse are essential to mitigate biases, ensure user privacy, and enhance transparency and accountability. This section discusses the proposal of ethical guidelines and best practices for AI in the Metaverse, including recommendations for addressing biases and discrimination, safeguarding user privacy and data protection, and strategies for enhancing transparency, explainability, and accountability in AI systems.

4.1 Proposal of ethical guidelines and best practices to mitigate biases and discrimination

This research seeks to address biases and discrimination in AI-driven virtual environments, safeguard user privacy and data protection, and enhance transparency, explainability, and accountability by proposing 10 ethical guidelines:

1. AI developers and platform operators should prioritise the use of diverse and representative training datasets [22]. This practice ensures that AI algorithms learn from a wide range of data, avoiding underrepresentation or marginalisation of certain demographic groups. Additionally, comprehensive data representation helps mitigate biases and ensures fairness in AI decision-making processes [59].
2. Regular auditing and bias detection mechanisms should be implemented to continuously monitor AI algorithms' performance (Landers [47] AI systems should be evaluated for potential sources of bias, and corrective measures should be taken to rectify biases whenever detected. Transparency in the auditing pro-

cess can foster accountability and trust among users [83].

3. Users should be made aware of the presence of AI algorithms and their potential impact on their experiences within the virtual environment [54]. Providing users with control over certain AI-driven features, such as personalised content recommendations, allows them to tailor their experiences and ensures that they are not subjected to biased or discriminatory content (Burrell [14].
4. AI systems within the Metaverse should be designed to be explainable. Users should have access to clear and understandable explanations of AI-driven decisions [63]. This transparency empowers users to challenge or question AI outcomes and fosters trust in AI technologies [59].
5. AI developers and platform operators should actively monitor AI systems for discriminatory outcomes [75]. If discrimination is identified, corrective actions should be taken promptly to address the issue and prevent its recurrence. Regular assessments of AI systems' fairness are crucial to ensure that they do not contribute to discriminatory practices [83].
6. Ethical guidelines should encourage interdisciplinary collaboration among AI developers, ethicists, policymakers, and representatives from impacted communities [45]. This diverse perspective allows for a comprehensive understanding of ethical challenges and fosters the development of contextually relevant solutions [28].
7. Before deploying AI systems in virtual environments, comprehensive ethical impact assessments should be conducted [55]. These assessments evaluate the potential ethical implications of AI use and inform decision-making processes. Ethical impact assessments contribute to responsible AI development and deployment [59].
8. Implementing feedback mechanisms for users can help identify and address potential biases or discrimination in AI systems [58]. Users should be encouraged to provide feedback on AI-driven experiences, and platforms should have mechanisms to consider and act on user feedback [14].
9. AI developers, platform operators, and other stakeholders involved in the Metaverse should receive continuous training on ethics in AI development and deployment [13]. Education on ethical considerations ensures that the AI community is well-equipped to address potential biases and discrimination [59].
10. To enhance accountability and transparency, collaboration with external auditors or independent organisations can be considered. External auditing can provide an objective assessment of AI systems and ensure that ethical guidelines are being followed [83].

Overall, the proposed ethical guidelines and best practices aim to mitigate biases and discrimination in AI-driven virtual environments. By prioritising fairness, transparency, and user empowerment, these guidelines can contribute to a more ethical and inclusive Metaverse experience for all users. Regular monitoring, user awareness, and collaboration among stakeholders are essential to uphold ethical standards in AI technologies within the virtual world.

4.2 Recommendations for ensuring user privacy and data protection in virtual worlds

Ensuring user privacy and data protection is paramount in the virtual worlds of the Metaverse, where AI technologies collect and analyse vast amounts of user data. To safeguard user privacy and maintain trust within virtual environments, the following recommendations are essential:

1. Adopt privacy-by-design principles in the development of AI systems and virtual environments [17]. This approach involves integrating privacy protections into the very design and architecture of AI applications, ensuring that privacy is considered from the outset rather than retrofitted as an afterthought [87].
2. Provide users with clear and accessible options to control the collection and use of their personal data. Users should be able to give informed consent for data processing and be empowered to modify or revoke their consent at any time [20].
3. Apply data anonymisation techniques to minimise the amount of personally identifiable information collected and stored [62]. Implement data minimisation practices to only collect the data necessary for specific AI functionalities within the virtual environment [14].
4. Implement robust security measures to protect user data from unauthorised access, breaches, and cyber-attacks [67]. Encryption and secure data storage mechanisms are essential to prevent data leaks (European [30].
5. Be transparent with users about data collection and usage practices. Provide clear and comprehensible explanations of how AI algorithms operate and handle user data. Educate users about the potential risks and benefits of AI in the virtual world to foster informed decision-making [59].
6. Conduct regular audits and impact assessments of AI systems' data practices to identify and mitigate potential privacy risks and vulnerabilities. This process should involve external experts and organisations to ensure objectivity [83].

7. Adhere to relevant data protection and privacy laws, such as the GDPR in the EU, the Data Security Law in China, and sector-specific laws in other jurisdictions.
8. Educate users about their rights and provide them with user-friendly tools to exercise their privacy preferences and access their personal data (Tene & Polonetsky [85].

By implementing these recommendations, virtual worlds can foster an environment where user privacy is respected, data is responsibly managed, and users feel empowered and confident in their interactions with AI systems.

4.3 Strategies for enhancing transparency, explainability, and accountability in AI systems within the Metaverse

Enhancing transparency, explainability, and accountability in AI systems within the Metaverse is essential to build user trust, ensure fairness, and foster responsible AI practices. The following strategies can be employed to achieve these goals:

1. Emphasise the use of explainable AI models within the Metaverse. AI algorithms should be designed to provide clear and understandable explanations for their decisions and recommendations [41]. Explainable AI not only helps users understand the rationale behind AI-driven actions but also enables developers and policymakers to identify and address potential biases [71].
2. Ensure transparency in the functioning of AI algorithms. Developers should make efforts to disclose information about the data sources, training processes, and decision-making criteria of AI systems to users [34]. Transparent algorithmic processes instill confidence and enable users to make informed choices within virtual environments.
3. Prioritise user-centric design when developing AI systems for the Metaverse. Taking user needs and preferences into account during the design phase can lead to more inclusive and personalised experiences (European [30]. User involvement in the development process can also promote acceptance and user satisfaction.
4. Implement external auditing and oversight mechanisms to evaluate AI systems' behavior within the virtual world [83]. Independent audits can help identify potential ethical concerns, biases, and areas for improvement. Additionally, external oversight bodies can ensure adherence to ethical guidelines and best practices.
5. Conduct regular algorithmic impact assessments to evaluate the potential social and ethical consequences of AI algorithms within the Metaverse [59]. Assessments should include the identification of potential biases, discrimination, and any unintended consequences of AI-driven decisions.
6. Establish user feedback mechanisms that allow users to report concerns and provide feedback on their AI experiences [14]. Prompt and effective redress mechanisms should be in place to address user complaints related to AI algorithm behavior.
7. Form ethics review boards or committees to oversee AI development and deployment within the Metaverse [9]. These boards can ensure that AI systems align with ethical guidelines and principles, and they can provide guidance on potential ethical challenges.
8. Foster collaboration among AI researchers, ethicists, policymakers, and stakeholders in the development of AI systems (European [30]. Multidisciplinary perspectives can contribute to a comprehensive understanding of ethical implications and lead to more robust and responsible AI practices (Brundage et al. [11].

By incorporating these strategies, developers and operators can promote transparency, explainability, and accountability in AI systems within the Metaverse, thus creating a more trustworthy and ethically sound virtual environment for users.

5 Case studies and ethical challenges

Following the discussions in previous chapters on ethical challenges and recommendations, this chapter delves into real-world case studies that shed light on the ethical implications of AI in virtual environments. Through a comprehensive exploration of these case studies, we aim to understand the complexities and potential pitfalls of AI implementation in the Metaverse.

5.1 Decentraland

Decentraland is a blockchain-based virtual world that allows users to own, create, and monetise content [24]. The platform relies on decentralised technology to enable user-driven interactions and transactions. However, Decentraland faces ethical challenges related to governance and content moderation [84]. As AI algorithms play a role in content curation and moderation, ensuring fair representation, diversity, and responsible AI practices is vital. The case of Decentraland highlights the need for robust governance mechanisms and transparent AI systems that foster inclusivity and protect user rights.

One of the prominent ethical concerns in Decentraland revolves around content moderation and ensuring age-appropriate experiences. As a user-generated content platform, Decentraland relies on AI algorithms for content curation and moderation to prevent the dissemination of inappropriate or harmful material [2]. However, determining what

content is suitable for different age groups and ensuring the protection of younger users can be challenging for AI systems [93].

To address this, Decentraland must implement robust content moderation policies and mechanisms that align with community standards and legal requirements. Combining AI-driven content filtering with community-driven reporting and human review can strike a balance between safety and creative freedom [50]. Moreover, Decentraland could provide users with better control over their interactions, allowing individuals to set their preferences and customise their experiences based on their comfort levels.

Another significant ethical consideration in Decentraland relates to governance and decision-making within the virtual world. As a decentralised platform, Decentraland operates through a Decentralised Autonomous Organisation (DAO), where decisions about policies and platform changes are made by token holders. This governance structure raises questions about representation, inclusivity, and the potential for concentration of power in the hands of a few stakeholders [77, 94].

To address these concerns, Decentraland should adopt transparent governance mechanisms that encourage participation from a diverse range of users. Encouraging broader token distribution and active participation from the community can promote inclusivity and prevent decision-making monopolies. Additionally, conducting regular audits of governance processes can help ensure accountability and fairness in decision-making.

Moreover, ensuring diverse and inclusive representation in the development and governance of Decentraland can foster a platform that reflects the values and preferences of its diverse user base. Collaborating with experts in ethics, diversity, and inclusion can provide valuable insights and help shape responsible AI practices that prioritise fairness and equitable representation.

5.2 Roblox

Roblox is a massively multiplayer online game creation platform that allows users to create and share their games and experiences [40]. With a large user base, Roblox presents ethical challenges concerning content moderation, safety measures, and age-appropriate experiences. AI algorithms are employed to moderate content and protect users from inappropriate or harmful material [21]. The case of Roblox illustrates the importance of developing AI-driven moderation systems that strike a balance between safety and creative freedom, ensuring a positive and secure virtual environment for users [74].

As in previous cases, one of the significant ethical concerns in Roblox is content moderation. As a user-generated content platform, Roblox relies on AI algorithms to automatically detect and filter inappropriate or harmful content [21]. However, striking the right balance between allowing creative freedom and ensuring a safe and positive experience for all users can be complex. AI-driven content moderation must be robust enough to prevent the dissemination of harmful content, such as violence or explicit material, while avoiding overly restrictive measures that stifle creativity [53].

To address this, Roblox must continue refining its AI content moderation systems, employing a combination of machine learning algorithms and human oversight. Human moderation teams can review flagged content, addressing potential false positives and nuances that AI may miss [15]. Additionally, Roblox could implement community-driven reporting mechanisms, enabling users to report inappropriate content, thereby empowering the community to actively participate in maintaining a safe virtual environment [82].

Another ethical challenge for Roblox is ensuring age-appropriate experiences for younger users. The platform attracts a substantial number of children and teenagers, necessitating measures to protect them from potentially harmful interactions or content unsuitable for their age group. Existing parent mode controls may be insufficient in addressing all risks [4]. AI algorithms play a critical role in age verification and ensuring that certain experiences are accessible only to users of appropriate age.

To enhance the age-appropriateness of experiences, Roblox must invest in AI-driven age verification systems that are accurate and reliable [51]. This could involve utilising machine learning techniques to analyse user behavior, interactions, and communication to assess their age more accurately. Ensuring robust age verification can create a safer space for younger users and build trust among parents and guardians.

5.3 Sansar

Sansar, developed by Wookey Project Corp, is a virtual social platform utilising AI-driven avatar customisation and interaction suggestions [76]. Ethical concerns encompass obtaining user consent for avatar data usage and addressing the potential impact of AI suggestions on user behavior and social dynamics. Transparent communication about data handling and designing AI systems to promote inclusivity and diverse interactions are essential for ensuring responsible AI use in the platform.

One of the primary ethical challenges in Sansar is obtaining explicit user consent for avatar data usage. As users interact and customise their virtual avatars, AI algorithms collect data to personalise their experiences and interactions within the platform. However, users may not always be fully aware of the extent to which their avatar data is being utilised or shared with third parties [86].

To address this, Sansar should implement clear and transparent communication about data handling and usage. Users should be provided with easy-to-understand privacy policies and consent mechanisms that explicitly outline how their avatar data will be used and whether it will be shared with other users or external entities. Additionally, Sansar should provide users with options to customise their data sharing preferences, allowing them to control the extent to which their avatar data is utilised for personalisation.

Another ethical concern in Sansar revolves around the potential impact of AI-driven interaction suggestions on user behavior and social dynamics. AI algorithms analyse user interactions and preferences to provide suggestions for virtual social interactions [70]. While this can enhance user experiences, there is a risk of reinforcing existing biases and creating echo chambers within the virtual world [81].

To mitigate this, Sansar must prioritise inclusive and diverse interaction suggestions. This involves implementing AI systems that promote diverse user experiences and social interactions, avoiding reinforcing stereotypes or segregating users into homogenous groups [90]. Additionally, Sansar can adopt user-centric design principles and involve user feedback to ensure that AI-driven suggestions are sensitive to user preferences while fostering inclusivity.

Furthermore, Sansar can encourage users to actively participate in shaping their AI-driven experiences. By providing users with the ability to provide feedback and rate the relevance and appropriateness of AI suggestions, the platform can continuously improve its AI systems and make them more responsive to users' needs and preferences [52].

6 Conclusion

In conclusion, the rise of AI in the Metaverse presents exciting opportunities for human interaction and creativity, but it also poses significant ethical challenges that demand thoughtful solutions. Through a study of real-world case studies, we have explored the complexities of AI implementation in virtual environments.

Key ethical concerns revolve around responsible content moderation, safeguarding user data privacy, and promoting inclusivity. Striking a balance between freedom of expression and user safety is crucial in AI-driven content curation, and human-AI hybrid moderation approaches show promise in achieving this balance. Additionally, transparent data handling, explicit user consent, and privacy by design principles are essential in protecting user privacy.

Furthermore, the governance of virtual worlds demands inclusive decision-making mechanisms to ensure representation and prevent concentration of power. Collaboration among platform operators, AI developers, ethicists, and users is vital to foster a responsible and inclusive Metaverse. By prioritising user well-being and employing responsible AI practices, we can shape the Metaverse into a transformative space that enriches lives while upholding ethical principles. As we navigate this evolving digital landscape, ethical considerations must remain at the forefront of our efforts to create a Metaverse that serves the collective good.

Author contributions The author conducted a comprehensive analysis of the ethical implications of AI in the Metaverse, including extensive research, synthesising information, structuring the paper, and revising the manuscript for clarity and accuracy.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. The author declares that no funding was received for this paper.

Data availability All data used in this article is based on publicly available information and can be accessed through the references cited. No new primary data was generated for this paper.

Declarations

Conflict of interest The author declares no competing interests regarding the publication of this article.

References

1. Adams, D., Bah, A., Barwulor, C., Musaby, N., Pitkin, K., & Redmiles, E. M. (2018). Ethics emerging: the story of privacy and security perceptions in virtual reality. In Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018) (pp. 427–442).
2. Age Rating & Scene Reporting. (n.d.). Decentraland Documentation. Retrieved February 14, 2024, from https://docs.decentraland.org/player/general/in-world-features/age-rating-scene-reporting/
3. Ahmet, E.F.E.: The impact of artificial intelligence on social problems and solutions: an analysis on the context of digital divide and exploitation. Yeni Medya **2022**(13), 247–264 (2022)
4. Allowed Experiences Controls. (n.d.). Roblox Help Center. Retrieved Retrieved February 14, 2024, from https://en.help.roblox.com/hc/en-us/articles/8863284850196-Allowed-Experiences-Controls#:~:text=Parents%20are%20able%20to%20select,below%20the%20age%20recommendations%20set
5. Balasubramaniam, N., Kauppinen, M., Hiekkanen, K., Kujala, S.: Transparency and explainability of AI systems ethical guidelines in practice. In: International working conference on requirements engineering: foundation for software quality, pp. 3–18. Springer International Publishing, Cham (2022)

6. Bang, J., Kim, J.Y.: Metaverse ethics for healthcare using AI technology challenges and risks. In: Rauterberg, M. (ed.) Culture and Computing HCII 2023 Lecture Notes in Computer Science, pp. 367–378. Springer, Cham (2023)
7. Bavana, K.: Privacy in the metaverse. Jus Corpus LJ **2**, 1 (2021)
8. Benjamins, R., Rubio Viñuela, Y., Alonso, C.: Social and ethical challenges of the metaverse: opening the debate. AI Ethics (2023). https://doi.org/10.1007/s43681-023-00278-5
9. Bernstein, M. S., Levi, M., Magnus, D., Rajala, B., Satz, D., & Waeiss, C. (2021). Esr: Ethics and society review of artificial intelligence research. arXiv preprint arXiv:2106.11521.
10. Bozkir, E., Özdel, S., Wang, M., David-John, B., Gao, H., Butler, K., Kasneci, E.: Eye-tracked virtual reality: a comprehensive survey on methods and privacy challenges. arXiv Prepr (2023). https://doi.org/10.48550/arXiv.2305.14080
11. Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., Anderljung, M.: Toward trustworthy AI development: mechanisms for supporting verifiable claims. arXiv Prepr (2020). https://doi.org/10.48550/arXiv.2004.07213
12. Buolamwini, J., Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77–91). PMLR
13. Burr, C., Leslie, D.: Ethical assurance: a practical approach to the responsible design, development, and deployment of data-driven technologies. AI Ethics **3**(1), 73–98 (2023)
14. Burrell, J.: How the machine 'thinks': understanding opacity in machine learning algorithms. Big Data Soc. (2016). https://doi.org/10.1177/2053951715622512
15. Calleberg, E. (2021). Making Content Moderation Less Frustrating: How Do Users Experience Explanatory Human and AI Moderation Messages
16. Cath, C.: Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Phil. Trans. R. Soc. A **376**(2133), 20180080 (2018)
17. Cavoukian, A.: Privacy by design and the promise of smartdata. In: SmartData: privacy meets evolutionary robotics, pp. 1–9. Springer, New York (2013)
18. Chen, P., Wu, L., Wang, L.: AI fairness in data management and analytics: a review on challenges. Methodol. Appl. Appl. Sci. **13**(18), 10258 (2023)
19. Cihon, P. (2019). Standards for AI governance: international standards to enable global coordination in AI research & development. Future of Humanity Institute. University of Oxford, 340–342.
20. Cohen, I.G.: Informed consent and medical artificial intelligence: What to tell the patient? Geo. LJ **108**, 1425 (2019)
21. Content Moderation on Roblox. (n.d.). Roblox Help Center. Retrieved February 14, 2024, from https://en.help.roblox.com/hc/en-us/articles/21416271342868-Content-Moderation-on-Roblox
22. Daugherty, P. R., Wilson, H. J., & Chowdhury, R. (2020). Using artificial intelligence to promote diversity. In: How AI Is Transforming the Organization, pp. 15-22
23. de Almeida, P.G.R., dos Santos, C.D., Farias, J.S.: Artificial intelligence regulation: a framework for governance. Ethics Inf. Technol. **23**(3), 505–525 (2021)
24. Decentraland. (n.d.). Decentraland White paper. Retrieved February 14, 2024, from https://decentraland.org/whitepaper.pdf
25. Di Pietro, R., Cresci, S. (2021, December). Metaverse: security and privacy issues. In 2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA) (pp. 281–288). IEEE
26. Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M.L., Herrera-Viedma, E., Herrera, F.: Connecting the dots in trustworthy artificial intelligence: from AI principles, ethics, and key requirements to responsible AI systems and regulation. Inform Fusion **99**, 101896 (2023)
27. Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Wood, A. (2017). Accountability of AI under the law: The role of explanation. arXiv preprint arXiv:1711.01134.
28. Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Williams, M.D.: Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int. J. Inform. Manage. **57**, 101994 (2021)
29. European Commission. (2019). Ethics Guidelines for Trustworthy AI. Retrieved on July 20, 2023, from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
30. European Parliament. (2016). 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) (Text with EEA relevance). Official Journal of the European Union, L 119, 1–88. Retrieved on July 20, 2023, from http://data.europa.eu/eli/reg/2016/679/oj
31. Fairfield, J.A.: Mixed reality: how the laws of virtual worlds govern everyday life. Berkeley Tech. LJ **27**, 55 (2012)
32. Falco, G., Shneiderman, B., Badger, J., Carrier, R., Dahbura, A., Danks, D., Yeong, Z.K.: Governing AI safety through independent audits. Nat. Mach. Intell. **3**(7), 566–571 (2021)
33. Federal Trade Commission. (n.d.). Federal Trade Commission Act. Retrieved on July 20, 2023, from https://www.ftc.gov/legal-library/browse/statutes/federal-trade-commission-act
34. Felzmann, H., Fosch-Villaronga, E., Lutz, C., Tamò-Larrieux, A.: Towards transparency by design for artificial intelligence. Sci. Eng. Ethics **26**(6), 3333–3361 (2020). https://doi.org/10.1007/s11948-020-00276-4
35. Ferrer, X., van Nuenen, T., Such, J.M., Coté, M., Criado, N.: Bias and discrimination in AI: a cross-disciplinary perspective. IEEE Technol. Soc. Mag. **40**(2), 72–80 (2021)
36. Fu, R., Huang, Y., & Singh, P. V. (2020). Artificial intelligence and algorithmic bias: Source, detection, mitigation, and implications. In Pushing the Boundaries: Frontiers in Impactful OR/OM Research (pp. 39–63). INFORMS
37. Global Partnership on Artificial Intelligence (GPAI). (n.d.). Retrieved on July 20, 2023, from https://www.gpai.ai/
38. Golbin, I., Rao, A. S., Hadjarian, A., & Krittman, D. (2020, December). Responsible AI: a primer for the legal community. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 2121–2126). IEEE
39. Habbal, A., Ali, M.K., Abuzaraida, M.A.: Artificial intelligence trust, risk and security management (AI TRiSM): frameworks, applications, challenges and future research directions. Expert Syst. Appl. **240**, 122442 (2024)
40. Hernández, L., Hernández, V., Neyra, F., Carrillo, J.: The use of massive online games in game-based learning activities. Rev. Innov. Educ. **4**(3), 7–30 (2022)
41. Hoffman, R.R., Mueller, S.T., Klein, G., Litman, J.: Measures for explainable AI: explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. Front. Comput. Sci. **5**, 1096257 (2023). https://doi.org/10.3389/fcomp.2023.1096257
42. Huang, Y., Li, Y.J., Cai, Z.: Security and privacy in metaverse: a comprehensive survey. Big Data Min. Anal. **6**(2), 234–247 (2023). https://doi.org/10.26599/BDMA.2022.9020047
43. Huynh-The, T., Pham, Q.V., Pham, X.Q., Nguyen, T.T., Han, Z., Kim, D.S.: Artificial intelligence for the metaverse: a survey. Eng. Appl. Artif. Intell. **117**, 105581 (2023)

44. Kaufmann, J., Hilgert, F., Wohlthat, R.: The proposed american data privacy and protection act in comparison with GDPR: does the current US bill of the ADPPA converge towards the "gold standard" concepts under the EU GDPR–or not? Comput. Law Rev. Int. **23**(5), 146–152 (2022)
45. Kusters, R., Misevic, D., Berry, H., Cully, A., Le Cunff, Y., Dandoy, L., Wehbi, F.: Interdisciplinary research in artificial intelligence: challenges and opportunities. Front Big Data **3**, 577974 (2020)
46. Lammerding, L., Hilken, T., Mahr, D., Heller, J.: Too real for comfort measuring consumers augmented reality information privacy concerns. In: Jung, T.H., et al. (eds.) Augmented reality and virtual reality progress, pp. 95–108. Springer, Cham (2021)
47. Landers, R.N., Behrend, T.S.: Auditing the AI auditors: a framework for evaluating fairness and bias in high stakes AI predictive models. Am. Psychol. **78**(1), 36 (2023)
48. Lee, M.K.: Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management. Big Data Soc. **5**(1), 2053951718756684 (2018)
49. Lee, N.T., Resnick, P., Barton, G.: Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms, p. 2. Brookings Institute, Washington, DC, USA (2019)
50. Li, Haoyan and Chau, Michael, "Human-AI Collaboration in Content Moderation: The Effects of Information Cues and Time Constraints" (2023). ECIS 2023 Research-in-Progress Papers. 2. https://aisel.aisnet.org/ecis2023_rip/2
51. Liu, L., Xiong, C., Zhang, H., Niu, Z., Wang, M., Yan, S.: Deep aging face verification with large gaps. IEEE Trans. Multim. **18**(1), 64–75 (2015)
52. Liu, S., Wright, A.P., Patterson, B.L., Wanderer, J.P., Turer, R.W., Nelson, S.D., Wright, A.: Using AI-generated suggestions from ChatGPT to optimize clinical decision support. J. Am. Med. Inform. Assoc. **30**(7), 1237–1245 (2023)
53. Llansó, E., van Hoboken, J., Leerssen, P., Harambam, J. (2020). Content Moderation, and Freedom of Expression. Algorithms
54. Luck, M., Aylett, R.: Applying artificial intelligence to virtual reality: intelligent virtual environments. Appl. Artif. Intell. **14**(1), 3–32 (2000)
55. Madary, M., Metzinger, T.K.: Real virtuality: a code of ethical conduct recommendations for good scientific practice and the consumers of VR-technology. Front Robot AI **3**, 3 (2016)
56. Maloney, D., Freeman, G., Robb, A. (2021, March). Social virtual reality: ethical considerations and future directions for an emerging research space. In 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 271–277). IEEE
57. McStay, A.: The metaverse: surveillant physics, virtual realist governance, and the missing commons. Philos. Technol. **36**(1), 13 (2023)
58. Meurisch, C., Mihale-Wilson, C.A., Hawlitschek, A., Giger, F., Müller, F., Hinz, O., Mühlhäuser, M.: Exploring user expectations of proactive AI systems. Proceed. ACM Interact. Mob. Wearable Ubiquitous Technol. **4**(4), 1–22 (2020)
59. Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L.: The ethics of algorithms: mapping the debate. Big Data Soc. (2016). https://doi.org/10.1177/2053951716679679
60. Moneta, A.: Architecture, heritage, and the metaverse: new approaches and methods for the digital built environment. Tradit. Dwell. Settl. Rev. **32**(1), 37–49 (2020)
61. Mystakidis, S.: Metaverse. Encyclopedia **2**(1), 486–497 (2022). https://doi.org/10.3390/encyclopedia2010031
62. Nair, V., Guo, W., O'Brien, J. F., Rosenberg, L., Song, D. (2023). Deep Motion Masking for Secure, Usable, and Scalable Real-Time Anonymization of Virtual Reality Motion Data. arXiv preprint arXiv:2311.05090.
63. Nassar, A., Kamal, M.: Ethical dilemmas in AI-powered decision-making: a deep dive into big data-driven ethical considerations. Int. J. Responsib. Artif. Intell. **11**(8), 1–11 (2021)
64. National People's Congress of the People's Republic of China. (2021). Data Security Law of the People's Republic of China. Retrieved on July 20, 2023, from http://www.npc.gov.cn/englishnpc/c23934/202112/1abd8829788946ecab270e469b13c39c.shtm
65. O'Brolcháin, F., Jacquemard, T., Monaghan, D., et al.: The convergence of virtual reality and social networks: threats to privacy and autonomy. Sci. Eng. Ethics **22**, 1–29 (2016). https://doi.org/10.1007/s11948-014-9621-1
66. O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., Ashrafian, H.: Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. Int. J. Med. Robot. Comput. Assist. Surg. **15**(1), e1968 (2019)
67. Odeleye, B., Loukas, G., Heartfield, R., Sakellari, G., Panaousis, E., Spyridonis, F.: Virtually secure: a taxonomic assessment of cybersecurity challenges in virtual reality environments. Comput. Secur. **124**, 102951 (2023)
68. Ooi, B. C., Chen, G., Shou, M. Z., Tan, K. L., Tung, A., Xiao, X., Zhang, M. (2022). The Metaverse Data Deluge: What Can We Do About It?. arXiv preprint arXiv:2206.10326.
69. Parlar, T.: Data Privacy and Security in the Metaverse. In: Metaverse: Technologies, pp. 123–133. Opportunities and Threats. Singapore, Springer Nature Singapore (2023)
70. Privacy Policy. (n.d.). Sansar. Retrieved February 14, 2024, from https://docs.sansar.com/untitled/guidelinesmoderation/guidelines-and-policies/privacy-policy#how-do-we-use-the-information-we-obtain
71. Rai, A.: Explainable AI: from black box to glass box. J. Acad. Mark. Sci. **48**, 137–141 (2020)
72. 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) (Text with EEA relevance). (2016). Official Journal of the European Union, L 119(1)
73. Ribeiro, M. T., Singh, S., Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. https://doi.org/10.1145/2939672.2939778. pp. 1135–1144
74. Roblox. (n.d.). Roblox: Imagine, create, and play together. Retrieved on July 20, 2023, from https://www.roblox.com
75. Roselli, D., Matthews, J., & Talagala, N. (2019, May). Managing bias in AI. In Companion Proceedings of The 2019 World Wide Web Conference (pp. 539–544).
76. Sansar. (n.d.). Sansar: Virtual Worlds, Avatars, Free 3D Chat. Retrieved on July 20, 2023, from https://www.sansar.com
77. Santana, C., Albareda, L.: Blockchain and the emergence of Decentralized Autonomous Organizations (DAOs): An integrative model and research agenda. Technol. Forecast. Soc. Chang. **182**, 121806 (2022)
78. Sap, M., Card, D., Gabriel, S., Choi, Y., Smith, N. A. (2019). The risk of racial bias in hate speech detection. In Proceedings of the 57th annual meeting of the association for computational linguistics (pp. 1668–1678). https://doi.org/10.18653/v1/P19-1163
79. Schiff, D., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2020). Principles to practices for responsible AI: closing the gap. arXiv preprint arXiv:2006.04707.
80. Schmitt, L.: Mapping global AI governance: a nascent regime in a fragmented landscape. AI Ethics **2**(2), 303–314 (2022)

81. Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., Hall, P.: Towards a standard for identifying and managing bias in artificial intelligence. NIST Spec. Publ. **10**, 6028 (2022)
82. Seering, J.: Reconsidering self-moderation: the role of research in supporting community-based models for online content moderation. Proceed. ACM Human-Comput. Interact. **4**(CSCW2), 1–28 (2020)
83. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the conference on fairness, accountability, and transparency (pp. 59–68). https://doi.org/10.1145/3287560.3287598
84. Tan, A. (2021). Metaverse Realities: A Journey Through Governance, Legal Complexities, and the Promise of Virtual Worlds. Legal Complexities, and the Promise of Virtual Worlds (December 9, 2021).
85. Tene, O., Polenetsky, J.: To track or" do not track": advancing transparency and individual control in online behavioral advertising. Minn. JL Sci. & Tech. **13**, 281 (2012)
86. Terms of Service. (n.d.). Sansar. Retrieved February 14, 2024, from https://www.sansar.com/terms-of-service
87. van Rest, J., Boonstra, D., Everts, M., van Rijn, M., & van Paassen, R. (2014). Designing privacy-by-design. In Privacy Technologies and Policy: First Annual Privacy Forum, APF 2012, Limassol, Cyprus, October 10–11, 2012, Revised Selected Papers 1 (pp. 55–72). Springer Berlin Heidelberg
88. Van Rijmenam, M. (2022). Step into the metaverse: How the immersive internet will unlock a trillion-dollar social economy. John Wiley & Sons
89. Varona, D., Suárez, J.L.: Discrimination, bias, fairness, and trustworthy AI. Appl. Sci. **12**(12), 5826 (2022)
90. Varsha, P.S.: How can we manage biases in artificial intelligence systems–a systematic literature review. Int. J. Inform. Manage Data Insights **3**(1), 100165 (2023)
91. Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In Proceedings of the 2018 chi conference on human factors in computing systems (pp. 1–14). https://doi.org/10.1145/3173574.3174014
92. Visconti, R.M.: From physical reality to the metaverse: a multilayer network valuation. Journal of Metaverse **2**(1), 16–22 (2022)
93. Waelen, R.A.: The struggle for recognition in the age of facial recognition technology. AI Ethics **3**(1), 215–222 (2023)
94. Wang, S., Ding, W., Li, J., Yuan, Y., Ouyang, L., Wang, F.Y.: Decentralized autonomous organizations: concept, model, and applications. IEEE Trans. Comput. Soc. Syst. **6**(5), 870–878 (2019)
95. Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V, Schwartz, O. (2018). AI now report 2018. New York: AI Now Institute at New York University. (pp. 1–62)