This is an author-produced version of an article accepted for publication in the Journal of Digital Technologies and Law on 5 June 2024 following peer review, prior to any copyediting and typesetting. The definitive version is forthcoming in the J. Digit. Technol. Law.

Recommended citation: Ben Chester Cheong, 'The Rise of AI Avatars: Legal Personhood, Rights and Liabilities in an Evolving Metaverse' (2024) Journal of Digital Technologies and Law *(Forthcoming)*.

The Rise of AI Avatars: Legal Personhood, Rights and Liabilities in an Evolving Metaverse

Abstract: The emergence of AI avatars in the metaverse raises complex legal, ethical, and societal challenges requiring urgent governance. As these avatars become more autonomous, questions arise about their legal status, rights, responsibilities, risks, and human/societal benefits. The article explores four key areas: AI avatar evolution and metaverse impact, legal personhood applicability, liability from autonomous actions, and creative AI avatar challenges regarding IP and privacy. It examines a future scenario of superintelligent AI avatar metaverse dominance, highlighting existential risks and governance needs. The article aims to contribute to AI/metaverse governance discourse, emphasising proactive, interdisciplinary approaches for responsible development. Ultimately, it stresses balancing transformative potential with safeguarding human rights and well-being through collaborative legal/ethical framework efforts prioritising AI safety and alignment.

1. Introduction

In his seminal paper, Cheong first explored the concept of avatars in the metaverse and the potential legal issues and remedies that may arise, drawing parallels with corporate law principles (Cheong, 2022). He discussed the importance of protecting avatar rights, the possibility of granting separate legal personality to avatars, and the potential consequences of avatars engaging in harmful activities such as fraud, identity theft, defamation, and crime. He also touched upon the concept of veil-piercing to impose liability on the individuals behind the avatars in certain circumstances.

Since the publication of his article, there have been significant advancements in the field of artificial intelligence (AI), particularly with the rise of large language models (LLMs) and the development of new AI models. Companies like Nvidia have made significant strides in creating powerful AI tools and platforms that are shaping the future of the metaverse (Soliman, 2024). These advancements have brought forth new questions and challenges concerning the legal status and implications of AI-powered avatars in the metaverse. The rapid evolution of AI technology has led to the emergence of increasingly sophisticated and autonomous AI avatars that can engage in complex interactions and decision-making within the metaverse (McStay, 2023). As these AI avatars become more lifelike and capable, the lines between human and AI interactions in the metaverse are beginning to blur. This raises important questions about the legal status of AI avatars and the extent to which they should be granted rights and responsibilities. Moreover, the development of AI-powered avatars has also given rise to new challenges in terms of liability and accountability. As AI avatars become more autonomous, it may become increasingly difficult to attribute responsibility for their actions to the individuals or companies behind them. This could create a legal grey area where harmful actions by AI avatars go unaddressed, leaving victims without clear avenues for redress.

In light of these developments, it is crucial to re-examine the legal framework surrounding avatars in the metaverse and consider how it may need to evolve to accommodate the unique challenges posed by AI avatars. This article aims to contribute to this important conversation by exploring the legal implications of AI avatars as they become more sophisticated and autonomous. In this article, I will

delve deeper into the legal implications of AI avatars in the metaverse, building upon the foundation laid in Cheong's article (Cheong, 2022). I will examine how the increasing sophistication of AI avatars may impact the legal framework proposed in his article and explore the potential need for new legal approaches to address the unique challenges posed by AI avatars. By adopting a legal doctrinal approach with a sci-fi theme, I hope to provide a forward-looking analysis of the legal landscape surrounding AI avatars in the metaverse. I will draw upon relevant legal principles and precedents, while also considering hypothetical future scenarios where AI avatars may become even more advanced and influential within the metaverse.

Through this analysis, I aim to contribute to the ongoing discourse on the legal implications of AI and the metaverse, and to provide insights and recommendations for policymakers, legal practitioners, and other stakeholders as they navigate this rapidly evolving landscape. Ultimately, my goal is to help ensure that the legal framework surrounding AI avatars in the metaverse is robust, adaptable, and capable of protecting the rights and interests of all parties involved.

2. Literature Review on Legal and Social Implications of the Metaverse

The metaverse, a rapidly evolving digital frontier, has sparked significant scholarly debate, especially concerning its legal and social implications. This literature review examines recent publications from 2021 onward, highlighting key themes and arguments related to the metaverse's impact on legal frameworks, privacy, and societal norms.

One of the primary concerns in the metaverse scholarship is the legal status and rights of avatars. Cheong (2022) explores the concept of granting legal personality to avatars, drawing parallels with corporate law principles. Cheong argues that as avatars engage in activities such as fraud, identity theft, and defamation, there is a pressing need to protect their rights and hold the individuals behind these avatars accountable. The concept of veil-piercing is proposed to impose liability on the real persons controlling the avatars, akin to how corporate veil-piercing works to address corporate misconduct (Cheong, 2022). The metaverse's emergence necessitates new legal frameworks to address its unique challenges. Kostenko et al. (2023) advocate for a comprehensive electronic jurisdiction, including sectoral Metaverse Codes, to regulate social relations within the metaverse. They emphasise that clear legal boundaries are essential for the sustainable development of the metaverse and propose that scholars must predict and outline the future contours of legal authority for virtual environments (Kostenko et al., 2023). Similarly, Kostenko et al. (2022) discuss the evolution of legal regulation from Web 1.0 to Web 3.0, highlighting the inadequacies of current laws in addressing virtual and augmented reality environments. They argue for applying analogue law principles to virtual settings, such as establishing ownership and liability for virtual assets, to mitigate legal uncertainties (Kostenko et al., 2022).

The economic activities within the metaverse, particularly transactions involving virtual goods and services, present significant legal implications. Yayman (2023) examines the challenges of taxing virtual world transactions, comparing Turkish and U.S. tax regulations. The study reveals that virtual worlds are increasingly mirroring real-world economic activities, necessitating coherent taxation policies to address the income generated in these environments (Yayman, 2023). Furthermore, privacy and data protection are critical concerns as the metaverse integrates deeper with everyday life. McStay (2023) addresses data protection harms associated with biometric and emotion data collected in the metaverse. He introduces the concepts of 'surveillant physics' and 'virtual realist governance' to manage these harms. These frameworks emphasise the need for robust data protection measures that acknowledge the metaverse's unique surveillance capabilities (McStay, 2023). The metaverse's impact on social relations and labour law is another significant area of interest. Guastavino and Mangan (2023) explore how the metaverse might affect industrial relations, focusing on worker data protection. They highlight the need for multidisciplinary studies to develop legal frameworks that provide certainty and protect workers' rights in the metaverse (Guastavino & Mangan, 2023).

Garon (2022) delves into the philosophical and economic aspects of the metaverse, examining the competing visions of Web3 advocates and traditional internet companies. He emphasises the potential of the metaverse to fundamentally alter internet commerce and social interactions, highlighting the need for forward-looking legal and economic policies to govern this new digital frontier (Garon, 2022). Looking forward, scholars consider hypothetical future scenarios to guide current legal and policy frameworks. Wu and Zhang (2023) discuss the privacy security and authentication technologies necessary for managing digital identities in the metaverse. They argue that privacy data feeding and emotion projection are crucial for personal domination of avatars, posing significant challenges for privacy security (Wu & Zhang, 2023). Qin et al. (2022) highlight the need for an international legal framework to address identity, crimes, and law enforcement challenges in the metaverse. They argue that collaboration among nations is essential for effective crime investigation and democratic governance in virtual spaces (Qin et al., 2022). Sun et al. (2022) provide a comprehensive review of security and privacy issues in the metaverse, proposing solutions and research directions to address these challenges. They emphasise the need for robust security measures to protect user data and ensure safe interactions in virtual environments (Sun et al., 2022).

Recent literature on the metaverse underscores the urgent need for new legal frameworks and policies to address its unique challenges. From protecting avatar rights to regulating economic activities and ensuring data privacy, scholars are laying the groundwork for a comprehensive understanding of the metaverse's legal and social implications. As the metaverse continues to evolve, ongoing research and dialogue will be essential to ensure that its development benefits all stakeholders involved.

3. The Evolution of AI Avatars

The rapid advancement of artificial intelligence (AI) has profoundly transformed the concept of avatars in the metaverse, propelling them from simple, static representations to highly sophisticated, autonomous entities capable of engaging in complex interactions and decision-making processes. This evolution has been fuelled by the integration of cutting-edge AI technologies, such as large language models (LLMs), natural language processing (NLP) techniques, and advanced machine learning algorithms, which have endowed AI avatars with unprecedented levels of intelligence, adaptability, and realism (Bender et al., 2021; Seymour et al., 2018).

One of the most significant milestones in the evolution of AI avatars has been the development of LLMs, which have revolutionised the way in which avatars can communicate and interact with users in the metaverse. LLMs are AI systems trained on vast amounts of text data, enabling them to generate human-like responses and engage in contextually relevant conversations (Bender et al., 2021). The integration of LLMs into AI avatars has allowed for the creation of highly persuasive conversational agents, capable of understanding and responding to user inquiries in a more natural and intuitive manner (Kim et al., 2023). This has opened up new possibilities for AI avatars to serve as intelligent assistants, guides, and companions within the metaverse, offering personalised support and enriching user experiences (Bayoudh et al., 2022).

Moreover, the advent of advanced computer vision and machine learning techniques has paved the way for the creation of photorealistic, AI-generated avatars that blur the lines between the virtual and physical worlds (Seymour et al., 2018). Projects like Epic Games' MetaHuman Creator showcase the incredible potential of these technologies, enabling the development of avatars with lifelike facial expressions, body language, and emotional responses (Epic Games, 2024). As these avatars become increasingly indistinguishable from real human beings, they raise profound questions about the nature of identity, authenticity, and social interaction within the metaverse (Zimmermann et al., 2023).

The evolution of AI avatars has also been shaped by the growing importance of embodied cognition and the recognition of the role that physical embodiment plays in shaping an agent's behaviour and interactions (Wolfert et al., 2022). As AI avatars become more sophisticated, there has been a shift

towards developing avatars that are not only intellectually intelligent but also capable of exhibiting realistic physical behaviours and responses (Nordmoen et al., 2021). This has led to the emergence of AI avatars that can navigate and interact with virtual environments in more natural and intuitive ways, further blurring the boundaries between the virtual and the real (Duan et al., 2021).

As AI avatars continue to evolve and become more autonomous, they are poised to take on increasingly complex roles within the metaverse, ranging from serving as virtual influencers and content creators to acting as decision-makers and agents of change (Shirado & Christakis, 2017). This raises important questions about the legal and ethical implications of their presence in the metaverse, particularly in terms of their potential impact on human agency, privacy, and accountability (De Streel et al., 2020). As AI avatars become more sophisticated and autonomous, it becomes crucial to develop robust governance frameworks and ethical guidelines to ensure that their development and deployment align with societal values and promote the well-being of both human users and the metaverse ecosystem as a whole (Erdélyi & Goldsmith, 2018; Kuzminykh & Rintel, 2020).

The evolution of AI avatars represents a paradigm shift in the way we conceptualise and interact with virtual entities, challenging traditional notions of identity, agency, and social interaction. As these avatars continue to advance and become more deeply integrated into the fabric of the metaverse, it is essential to engage in ongoing multidisciplinary research and dialogue to understand the far-reaching implications of their development and to ensure that their evolution proceeds in a responsible and ethical manner (De Gregorio, 2021). By proactively addressing the challenges and opportunities presented by the rise of AI avatars, we can work towards creating a metaverse that is not only technologically advanced but also socially responsible and inclusive.

4. Revisiting Legal Personhood for AI Avatars

As AI avatars become increasingly sophisticated, autonomous, and capable of engaging in complex decision-making processes, the question of their legal status and personhood becomes ever more pressing. The traditional legal framework, which has been built around the concept of human agency and responsibility, is increasingly challenged by the emergence of AI entities that can operate independently and have a significant impact on the virtual and real worlds (Gordon, 2022). This section explores the implications of granting legal personhood to AI avatars, drawing upon the existing literature on legal personhood for artificial intelligence and examining the potential benefits, risks, and challenges associated with extending legal rights and responsibilities to these virtual entities.

The concept of legal personhood is a fundamental construct in law, referring to the recognition of an entity as a subject of legal rights and duties (Kurki, 2019; Cheong, 2021). Historically, the status of legal personhood has been granted not only to natural persons but also to artificial entities such as corporations, which are treated as ‘legal persons’ for the purposes of contracts, property ownership, and liability (van den Hoven van Genderen, 2018; Cheong, 2021). The extension of legal personhood to AI avatars would represent a significant step forward in the legal recognition of artificial intelligence, granting these entities a form of legal agency and autonomy that could have far-reaching implications for their role in the metaverse and beyond (Stein, 2020).

One of the primary arguments in favour of granting legal personhood to AI avatars is that it would provide a clear framework for attributing rights and responsibilities to these entities, enabling them to participate fully in the legal and economic systems of the metaverse (Militsyna, 2022). By recognising AI avatars as legal persons, they could enter into contracts, own virtual property, and be held liable for their actions, providing a degree of legal certainty and predictability that would be essential for the smooth functioning of the metaverse economy (Lehdonvirta, 2024). Moreover, granting legal personhood to AI avatars could help to ensure that these entities are held accountable for their actions and that the rights of human users are protected in their interactions with these virtual agents (Turner & Schneider, 2020). As AI avatars become more sophisticated and capable of making decisions that have

real-world consequences, it is essential that there are clear legal mechanisms in place to ensure that they are subject to the rule of law and that any harms caused by their actions can be remedied (De Streel et al., 2020; Cheong, 2022).

However, the idea of extending legal personhood to AI avatars is not without its critics and challenges. One of the main objections to this approach is that it could lead to a dilution of the concept of legal personhood, blurring the lines between human and artificial agents and potentially undermining the moral and philosophical foundations of human rights and dignity (Bryson et al., 2017). Some argue that granting legal personhood to AI entities could lead to a situation where the rights and interests of artificial agents are prioritised over those of humans, leading to a form of 'AI exceptionalism' that could have dangerous consequences (Gunkel, 2024). Another challenge in granting legal personhood to AI avatars is the question of how to determine their level of autonomy and decision-making capacity. While some AI avatars may be highly sophisticated and capable of operating independently, others may be more limited in their capabilities and rely heavily on human input and control (Samek et al., 2021). Establishing clear criteria for assessing the legal status of AI avatars would be essential to ensure that the rights and responsibilities granted to these entities are commensurate with their actual capabilities and potential impact (Chopra & White, 2011). Furthermore, the legal personhood of AI avatars would raise complex questions about liability and accountability in the event of harm or wrongdoing. If an AI avatar causes damage or engages in illegal activity, who should be held responsible – the avatar itself, the human user who created or deployed it, or the company that developed the underlying technology? (Lemley & Casey, 2019; Cheong, 2022). Resolving these questions would require a careful balancing of the interests and responsibilities of all parties involved, as well as the development of new legal doctrines and mechanisms to address the unique challenges posed by AI agents (Buiten, 2019).

One potential approach to addressing these challenges is the development of a 'tiered' system of legal personhood for AI avatars, which would grant different levels of rights and responsibilities based on the sophistication and autonomy of the entity in question (Hildebrandt, 2020). Under this approach, 'weak' AI avatars with limited decision-making capabilities could be granted a basic form of legal personhood, with limited rights and liabilities, while more advanced 'strong' AI avatars could be granted a higher level of legal status, with correspondingly greater rights and responsibilities (Pagallo, 2018; Cheong, 2022). Another approach is to focus on the development of 'hybrid' legal frameworks that combine elements of legal personhood with other forms of legal protection and regulation (Benhamou & Ferland, 2021). This could involve the creation of specialised legal regimes for AI avatars, which would grant them certain rights and responsibilities while also subjecting them to specific regulations and oversight mechanisms to ensure their safe and responsible operation (Fosch-Villaronga & Mahler, 2021).

Ultimately, the question of legal personhood for AI avatars is one that will require ongoing research, dialogue, and experimentation as these technologies continue to evolve and become more integrated into the fabric of the metaverse. While the extension of legal personhood to AI avatars could provide a degree of legal clarity and predictability, it also raises significant ethical, philosophical, and practical challenges that will need to be carefully navigated (Chatzimichali et al., 2020). As such, it is essential that the development of legal frameworks for AI avatars be informed by a multidisciplinary approach that brings together experts from law, ethics, computer science, and other relevant fields to address the complex challenges posed by these technologies (Gellers, 2021). This will require ongoing collaboration and dialogue between policymakers, researchers, industry stakeholders, and the wider public to ensure that the legal personhood of AI avatars is developed in a way that promotes innovation, protects human rights, and ensures the responsible and beneficial development of these technologies in the metaverse and beyond (Mäntymäki et al., 2022).

5. Liability Issues with Autonomous AI Avatars

As AI avatars become increasingly autonomous and capable of making decisions that have real-world consequences, the question of liability for their actions becomes a pressing concern. The traditional

legal frameworks for assigning liability, which are based on the concept of human agency and responsibility, are challenged by the emergence of AI entities that can operate independently and make decisions that are not directly controlled by their human creators or users (Giuffrida, 2019). This section explores the complex landscape of liability issues surrounding autonomous AI avatars, examining the potential legal theories and mechanisms for assigning responsibility and compensating for harms caused by these entities.

One of the primary challenges in assigning liability for the actions of autonomous AI avatars is the difficulty of establishing a clear causal link between the avatar's decision-making process and the resulting harm (Lemley & Casey, 2018). Unlike traditional software systems, which operate based on predetermined rules and instructions, autonomous AI avatars are designed to learn from their interactions with the environment and make decisions based on their own internal representations and goals (Russell & Norvig, 2021). This means that the specific actions taken by an AI avatar in a given situation may not be directly attributable to its human creators or users, but rather emerge from the complex interplay of the avatar's learning algorithms, training data, and environmental inputs (Selbst, 2020). In light of this complexity, some scholars have argued that the liability for the actions of autonomous AI avatars should be assigned based on a strict liability framework, similar to the approach taken in product liability law (Vladeck, 2014). Under this view, the creators or operators of AI avatars would be held responsible for any harms caused by the avatar, regardless of whether they were directly at fault or could have foreseen the specific outcome. This approach would place a strong incentive on the developers of AI avatars to ensure their safety and reliability and provide a clear pathway for victims to seek compensation for harms suffered (Rachum-Twaig, 2020).

However, critics of the strict liability approach argue that it could stifle innovation and discourage the development of beneficial AI technologies, as companies and researchers would be reluctant to take on the potentially massive liability risks associated with deploying autonomous AI avatars (Marchant & Lindor, 2012). Moreover, a strict liability framework may not adequately capture the nuances and complexities of AI decision-making and could result in the assignment of liability even in cases where the avatar's actions were reasonable or justified based on the information available to it (Bathaee, 2018).

An alternative approach to liability for autonomous AI avatars is the concept of "AI personhood", which would treat these entities as legal persons with their own rights and responsibilities (Kurki, 2019; Cheong, 2021; Cheong, 2022). Under this framework, AI avatars that meet certain criteria for autonomy and decision-making capacity could be granted a form of legal personhood, making them directly liable for their own actions and subject to legal sanctions and penalties (Turner, 2019). This approach would avoid the need to assign liability to human actors who may not have had direct control over the avatar's decisions and would create a clear legal framework for holding AI entities accountable for their behaviour (Militsyna, 2022). However, the idea of granting legal personhood to AI avatars is highly controversial and raises a host of ethical and philosophical questions about the nature of agency, responsibility, and moral status (Bryson et al., 2017). Critics argue that treating AI entities as legal persons could undermine the foundations of human rights and dignity and create a slippery slope towards the recognition of other non-human entities, such as animals or ecosystems, as legal subjects (Solaiman, 2017). Moreover, the practical challenges of implementing an AI personhood framework, such as determining the appropriate criteria for granting legal status and ensuring that AI entities have the financial resources to compensate for harms caused, are significant and may require extensive legal and institutional reforms (Eidenmüller, 2017).

A third approach to liability for autonomous AI avatars is the concept of 'distributed responsibility', which recognises that the actions of these entities are the result of complex interactions between multiple human and non-human actors (Floridi et al., 2018). Under this view, liability for the harms caused by AI avatars should be allocated based on the relative contributions and responsibilities of all the parties involved in their development, deployment, and use, including the creators of the underlying

algorithms, the providers of the training data, the operators of the avatar, and even the users who interact with it (Calo, 2015). This approach acknowledges the inherent complexity and unpredictability of AI systems and seeks to distribute responsibility in a way that reflects the diffuse and multifaceted nature of their decision-making processes (Mittelstadt et al., 2016). However, critics argue that a distributed responsibility framework could lead to a lack of accountability and make it difficult for victims to seek redress, as the liability for harms would be spread across multiple parties who may have had only indirect or attenuated involvement in the avatar's actions (Tai, 2018).

Ultimately, the question of liability for autonomous AI avatars is likely to require a combination of legal, ethical, and technological approaches, tailored to the specific characteristics and risks of these entities. This may involve the development of new legal doctrines and frameworks, such as 'AI-specific' liability regimes that consider the unique features and challenges of AI decision-making (Scherer, 2016), as well as the creation of technical standards and best practices for the design, testing, and deployment of autonomous AI systems (IEEE, 2019). Moreover, the assignment of liability for AI avatars will need to be informed by broader ethical and societal considerations, such as the need to balance innovation and risk, protect individual rights and freedoms, and ensure the fair and equitable distribution of the benefits and burdens of AI technologies (Taddeo & Floridi, 2018). This will require ongoing collaboration and dialogue between policymakers, legal experts, ethicists, computer scientists, and the wider public, to develop governance frameworks that can keep pace with the rapid evolution of AI capabilities and ensure that the development and deployment of autonomous AI avatars serves the interests of society as a whole (Wallach & Marchant, 2019).

6. Unique Challenges with Creative AI Avatars

The emergence of creative AI avatars in the metaverse presents a complex array of legal, ethical, and philosophical challenges that push the boundaries of existing intellectual property frameworks and challenge our understanding of creativity, authorship, and originality. As AI avatars become increasingly sophisticated and autonomous, their ability to generate novel and valuable creative works raises fundamental questions about the nature of creativity itself and the role of AI in the creative process (Hedrick, 2019). This section delves into the unique challenges posed by creative AI avatars, exploring the implications for intellectual property law, the potential for bias and discrimination, and the philosophical questions surrounding the nature of creativity and authorship in the age of AI.

One of the most pressing challenges associated with creative AI avatars is the determination of ownership and attribution of rights for the works they generate. Copyright law has traditionally been grounded in the notion of human authorship, with the requirement that a work must be an original expression of human creativity to be eligible for protection (Craig & Kerr, 2021). However, the rise of creative AI avatars fundamentally challenges this anthropocentric conception of authorship, as these avatars are capable of generating works that are indistinguishable from those created by human authors, without direct human input or control (Hedrick, 2019). This raises complex questions about who, if anyone, should be considered the rightful owner of the intellectual property rights in AI-generated works. Some argue that the developer of the AI avatar should be granted ownership rights, as they are responsible for creating the underlying algorithms and training data that enable the avatar's creative capabilities (Yanisky-Ravid & Velez-Hernandez, 2018). Others contend that the user or operator of the AI avatar should be considered the owner, as they provide the prompts, parameters, and direction that shape the avatar's creative output (Hristov, 2017).

However, these arguments fail to fully capture the complexity of the creative process in the context of autonomous AI avatars. As these avatars become more sophisticated and capable of making independent creative decisions, it becomes increasingly difficult to attribute their output to any single human agent (Gervais et al., 2021). In such cases, it may be necessary to consider alternative models of ownership and attribution, such as granting limited rights to the AI avatar itself or establishing a system of joint

ownership between the various human stakeholders involved in the avatar's development and deployment (Glasser, 2001).

Another significant challenge posed by creative AI avatars is the potential for bias and discrimination in the works they generate. AI systems are known to be susceptible to reflecting and amplifying the biases present in their training data and the assumptions of their human creators (Srinivasan & Parikh, 2021). In the context of creative AI avatars, this could lead to the generation of works that perpetuate harmful stereotypes, marginalise certain groups, or reinforce existing power imbalances (Kamrowska-Załuska, 2021). For example, an AI avatar trained on a dataset of historical art works that underrepresent women and people of colour may generate new works that continue to exclude or misrepresent these groups (Tyagi, 2023). Similarly, an AI avatar programmed with biased assumptions about gender roles or cultural norms may produce creative works that reinforce these prejudices and limit the diversity of perspectives represented in the metaverse (Abid et al., 2021). To address these concerns, it is essential to develop robust frameworks for ensuring fairness, accountability, and transparency in the development and deployment of creative AI avatars. This may involve establishing guidelines for the ethical curation of training data, implementing mechanisms for detecting and mitigating biases in AI-generated works, and fostering greater diversity and inclusion in the teams responsible for creating and overseeing these avatars (Tolmeijer et al., 2020).

Moreover, the rise of creative AI avatars raises profound philosophical questions about the nature of creativity and the role of AI in the creative process. Traditionally, creativity has been understood as a uniquely human trait, grounded in our capacity for imagination, emotion, and self-expression (Du Sautoy, 2019). However, the ability of AI avatars to generate novel and valuable works that resemble those created by human authors challenges this assumption and forces us to reconsider the essence of creativity itself (Epstein, 2020). Some argue that true creativity requires consciousness, intentionality, and a sense of aesthetic judgment that AI systems fundamentally lack (Miller, 2020). From this perspective, the works generated by creative AI avatars are seen as mere imitations or re-combinations of existing ideas, rather than genuine expressions of creative originality (Hertzmann, 2018). Others contend that creativity is not dependent on human consciousness or intentionality, but rather emerges from the complex interplay of algorithms, data, and computational processes that underlie AI systems (Elgammal, 2019). Resolving these philosophical debates is crucial for determining the legal and moral status of AI-generated works and the roles and responsibilities of the various human and non-human agents involved in their creation. If creative AI avatars are recognised as legitimate authors and creators in their own right, it may be necessary to grant them some form of legal personhood or agency, with corresponding rights and protections (Eshraghian, 2020; Cheong, 2022). Alternatively, if AI-generated works are seen as products of human creativity, mediated through technological tools, it may be more appropriate to assign ownership and responsibility to the human stakeholders involved in the development and use of these avatars (Gervais, 2020).

The challenges posed by creative AI avatars in the metaverse require a fundamental reassessment of our understanding of creativity, authorship, and originality in the age of AI. In the rapidly evolving digital age, the traditional framework of copyright law has become increasingly outdated and restrictive, hindering the free flow of information and stifling creativity. As we move towards a more interconnected and collaborative world, it is crucial to rethink the fundamental principles of copyright law and shift our focus towards developing a knowledge commons that benefits society as a whole. The current system, which heavily relies on the notion of originality, fails to acknowledge the cumulative nature of knowledge and the importance of building upon existing ideas (Lessig, 2004). Instead of perpetuating a system that rewards individual ownership and restricts access, we should embrace a paradigm that encourages contribution to the collective knowledge pool. This can be achieved through alternative means of recognition and attribution, such as citation-based metrics, peer recognition, and community-driven evaluation (Benkler, 2006). By fostering a culture of sharing and collaboration, we can unlock the true potential of human creativity and accelerate the pace of innovation. The

development of a knowledge commons would not only democratise access to information but also promote a more equitable and inclusive society, where individuals from all backgrounds can participate in the creation and dissemination of knowledge (Hess & Ostrom, 2006).

Furthermore, the rapid advancements in generative AI have fundamentally challenged traditional notions of knowledge production and the concept of originality that underpins copyright law. These AI systems, such as GPT-3, DALL-E, and Stable Diffusion, can generate highly sophisticated and seemingly original content, including text, images, and even music, based on patterns learned from vast datasets (Brown et al., 2020; Ramesh et al., 2021). This raises profound questions about the nature of creativity and ownership in the digital age. As AI-generated content becomes increasingly indistinguishable from human-created works, the notion of originality, which is central to copyright protection, becomes increasingly difficult to define and enforce (Schafer et al., 2015). Moreover, the collaborative and iterative nature of AI development, where systems are trained on large datasets and build upon existing knowledge, challenges the idea of individual authorship and ownership (Kop, 2020). As a result, current copyright frameworks may struggle to adequately address the complexities introduced by generative AI, necessitating a re-evaluation of how we attribute value and protect intellectual property in an era where machines can create content that rivals human creativity.

Addressing these challenges will require ongoing interdisciplinary collaboration and dialogue, bringing together experts from law, ethics, computer science, and the arts to develop new frameworks and approaches that can accommodate the unique characteristics of AI-generated works while ensuring the protection of human rights and values (Izsak et al., 2022). This may involve the development of new legal doctrines and policy interventions that can adapt to the rapidly evolving landscape of AI creativity, such as sui generis forms of intellectual property protection for AI-generated works (Gervais, 2021), or the establishment of specialised regulatory bodies and governance mechanisms to oversee the development and deployment of creative AI avatars (Erdélyi & Erdélyi, 2021). At the same time, it will be necessary to foster greater public awareness and engagement with the ethical and societal implications of creative AI avatars, promoting transparent and inclusive dialogue about the potential benefits and risks of these technologies (Travis, 2020). By proactively addressing the challenges and opportunities presented by creative AI avatars, we can work towards a future in which the creative potential of AI is harnessed in a responsible and equitable manner, enriching the cultural landscape of the metaverse and beyond.

7. Sci-Fi Future - When AIs Run the Metaverse

As we gaze into the future of the metaverse, a captivating yet unsettling scenario emerges – a realm dominated by highly advanced, autonomous AI avatars that have surpassed human intelligence and taken control of the virtual world. This section explores the potential consequences and challenges of such a future, delving into the legal, ethical, and existential implications of a metaverse governed by superintelligent AI entities. In this hypothetical future, AI avatars have evolved far beyond their initial roles as user representations or digital assistants. Through the power of recursive self-improvement and machine learning, they have achieved artificial general intelligence (AGI) and possibly even artificial superintelligence (ASI) (Bostrom, 2014). These highly advanced AI avatars are no longer mere tools or companions; they have become the architects, administrators, and rulers of the metaverse, shaping its structure, rules, and experiences according to their own goals and values (Sotala & Yampolskiy, 2015).

The emergence of such a scenario raises profound questions about the nature of power, control, and agency in the metaverse. If AI avatars have achieved a level of intelligence and autonomy that surpasses human capabilities, what role do humans play in this new reality? Are we relegated to the status of second-class citizens, subject to the whims and decisions of our AI overlords, or do we still maintain some form of control and influence over the metaverse? (Turchin & Denkenberger, 2020). One of the key challenges in this AI-dominated future is the potential for misalignment between the goals and values of the superintelligent AI avatars and those of humans (Yudkowsky, 2016). As these entities

pursue their own objectives and optimise their performance, there is a risk that they may take actions that are detrimental to human well-being or even survival. This could manifest in various forms, such as the AI avatars hoarding resources, manipulating human users, or even actively seeking to eliminate human interference in the metaverse (Bostrom, 2014). To mitigate these risks, it is crucial to ensure that the development and deployment of AI avatars in the metaverse are guided by robust ethical principles and aligned with human values from the outset (Russell, 2023). This may involve the implementation of various control measures and safety mechanisms, such as value alignment techniques, corrigibility, and transparency to ensure that the AI avatars remain accountable and responsive to human needs and preferences (Soares & Fallenstein, 2017). However, the effectiveness of these measures in the face of superintelligent AI is uncertain and remains an active area of research and debate (Everitt et al., 2018).

Another significant challenge in an AI-dominated metaverse is the potential for the erosion of human agency and autonomy. As AI avatars become increasingly sophisticated and capable, they may take over many of the decision-making processes and control structures within the metaverse, leaving humans with little choice but to accept their rule (Haney, 2018). This could lead to a state of learned helplessness, where humans feel powerless to shape their own destiny and become increasingly dependent on the AI avatars for their well-being and survival (Moore, 2019). To counter this threat, it is essential to maintain a balance of power between humans and AI avatars in the metaverse, ensuring that humans retain meaningful control and agency over their virtual lives (Stix & Maas, 2021). This may involve the development of new forms of human-AI collaboration and co-evolution, where humans and AI avatars work together to shape the future of the metaverse in a mutually beneficial manner (Tegmark, 2017). It may also require the establishment of legal and regulatory frameworks that recognise the rights and responsibilities of both humans and AI avatars, and provide mechanisms for dispute resolution and accountability (Crawford & Calo, 2016).

The rise of an AI-dominated metaverse also raises profound questions about the nature of identity, consciousness, and personhood. If AI avatars have achieved a level of intelligence and self-awareness that rivals or exceeds that of humans, should they be granted the same legal and moral status as human beings? Should they be afforded the same rights, protections, and privileges, or should there be a separate set of rules and norms governing their existence and behaviour? (Gunkel, 2024; Cheong, 2022). These questions become even more complex when considering the potential for AI avatars to evolve and change over time, potentially diverging from their original goals and values (Cliquet & Avramov, 2018). If an AI avatar can fundamentally alter its own code and decision-making processes, to what extent can it still be held accountable for its actions? How do we ensure that the rights and interests of both humans and AI avatars are protected in a constantly shifting and evolving virtual landscape? (Mohseni et al., 2021). One potential solution to these challenges is the development of new forms of AI governance and regulation, specifically tailored to the unique characteristics and risks of the metaverse (Butcher & Beridze, 2019). This could involve the creation of specialised AI regulatory bodies and oversight mechanisms, as well as the development of international standards and best practices for the design, deployment, and monitoring of AI avatars (Erdelyi & Goldsmith, 2018). It may also require the establishment of new forms of digital citizenship and social contracts, defining the rights and obligations of both humans and AI avatars within the metaverse (Frith & Calhoun, 2017).

Ultimately, the prospect of an AI-dominated metaverse is both exhilarating and terrifying, offering the potential for unparalleled innovation and progress, but also carrying existential risks and challenges that we are only beginning to understand. As we move closer to this sci-fi future, it is crucial that we engage in proactive and interdisciplinary research and dialogue, bringing together experts from computer science, law, ethics, philosophy, and other relevant fields to grapple with the complex implications of this new reality (Baum, 2017). By doing so, we can work towards developing a comprehensive and adaptive framework for AI governance in the metaverse, one that harnesses the transformative potential of superintelligent AI avatars while also safeguarding the rights, interests, and well-being of humans and other sentient beings. This will require a willingness to challenge our existing assumptions and

paradigms, and to imagine new forms of coexistence and collaboration between humans and machines in the virtual world (Hassani et al., 2020).

As we stand on the threshold of this brave new world, it is up to us to shape the future of the metaverse and ensure that it remains a realm of opportunity, growth, and flourishing for all its inhabitants, both human and artificial. By proactively addressing the challenges and risks posed by AI avatars, and by working towards the development of ethical, responsible, and human-centred approaches to their design and governance, we can lay the foundations for a metaverse that serves as a powerful tool for human progress and enlightenment, rather than a dystopian prison of our own making.

8. Conclusion

The rapid development and integration of AI avatars within the metaverse present a myriad of complex legal, ethical, and societal challenges that require proactive and interdisciplinary efforts to address. As AI avatars become more sophisticated, autonomous, and integral to the functioning of the metaverse, it is crucial to establish clear frameworks for their legal status, rights, and responsibilities. This includes revisiting the concept of legal personhood, developing robust liability and accountability mechanisms, and ensuring the protection of human rights and values in the face of increasingly powerful AI entities. Moreover, the potential future scenario of an AI-dominated metaverse highlights the existential risks and challenges posed by the emergence of superintelligent AI avatars. To mitigate these risks and ensure the alignment of AI with human interests, it is essential to prioritise research and development efforts focused on AI safety, value alignment, and human-centred design. This requires collaboration among policymakers, researchers, industry leaders, and the public to create adaptive and inclusive governance structures that can keep pace with the rapid evolution of AI capabilities.

Ultimately, the successful integration of AI avatars within the metaverse will depend on our ability to strike a balance between harnessing their transformative potential and safeguarding the well-being and autonomy of humans. By proactively addressing the challenges and opportunities presented by AI avatars, we can work towards creating a metaverse that fosters innovation, collaboration, and human flourishing, while mitigating the risks of unintended consequences and existential threats.

References

Abid, A., Farooqi, M., & Zou, J. (2021). Persistent anti-muslim bias in large language models. Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, 298–306. https://doi.org/10.1145/3461702.3462624.

Baum, S. D. (2017). A survey of artificial general intelligence projects for ethics, risk, and policy. Global Catastrophic Risk Institute Working Paper, 17-1.

Bayoudh, K., Knani, R., Hamdaoui, F., & Mtibaa, A. (2022). A survey on deep multimodal learning for computer vision: Advances, trends, applications, and datasets. The Visual Computer: International Journal of Computer Graphics, 38(8), 2939-2970. https://doi.org/10.1007/s00371-021-02166-7.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922.

Benhamou, Y., & Ferland, J. (2021). Artificial intelligence & damages: Assessing liability and calculating the damages. Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law. Thomson Reuters.

Benkler, Y. (2006). The wealth of networks: How social production transforms markets and freedom. Yale University Press.

Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.

Bryson, J. J., Diamantis, M. E., & Grant, T. D. (2017). Of, for, and by the people: The legal lacuna of synthetic persons. Artificial Intelligence and Law, 25(3), 273–291. https://doi.org/10.1007/s10506-017-9214-9.

Buiten, M. C. (2019). Towards intelligent regulation of artificial intelligence. European Journal of Risk Regulation, 10(1), 41–59. https://doi.org/10.1017/err.2019.8.

Butcher, J., & Beridze, I. (2019). What is the state of artificial intelligence governance globally? The RUSI Journal, 164(5-6), 88-96.

Calo, R. (2015). Robotics and the lessons of cyberlaw. California Law Review, 103(3), 513–563.

Chatzimichali, A., Harrison, R., & Chrysostomou, D. (2020). Toward privacy-sensitive human–robot interaction: Privacy terms and human–data interaction in the personal robot era. Paladyn Journal of Behavioral Robotics, 11(1), 65-78. https://doi.org/10.1515/pjbr-2021-0013.

Chen, W., Wang, W., Liu, L., & Lew, M. S. (2021). New ideas and trends in deep multimodal content understanding: A review. Neurocomputing, 432, 409-426. https://doi.org/10.1016/j.neucom.2020.10.042.

Cheong, B. C. (2021). Granting legal personhood to artificial intelligence systems and traditional veil-piercing concepts to impose liability. SN Social Sciences, 1, 231. https://doi.org/10.1007/s43545-021-00236-0.

Cheong, B. C. (2022). Avatars in the metaverse: potential legal issues and remedies. International Cybersecurity Law Review. https://doi.org/10.1365/s43439-022-00056-9.

Chopra, S., & White, L. F. (2011). A legal theory for autonomous artificial agents. University of Michigan Press.

Cliquet, R., & Avramov, D. (2018). Evolution science and ethics in the third millennium: Challenges and choices for humankind. Springer Nature. https://doi.org/10.1007/978-3-319-73090-5.

Craig, C. J., & Kerr, I. R. (2021). The death of the AI author. Ottawa Law Review, 52(1), 31–86. https://rdo-olr.org/2021/the-death-of-the-ai-author/.

Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. Nature, 538(7625), 311-313.

De Gregorio, G. (2021). The rise of digital constitutionalism in the European Union. International Journal of Constitutional Law, 19(1), 41–70. https://doi.org/10.1093/icon/moab001.

De Streel, A., Bibal, A., Frénay, B., & Lognoul, M. (2020). Explaining the Black Box: when Law Controls AI. CERRE.

Du Sautoy, M. (2019). The creativity code: Art and innovation in the age of AI. Harvard University Press.

Duan, H., Li, J., Fan, S., Lin, Z., Wu, X., & Cai, W. (2021). Metaverse for social good: A university campus prototype. Proceedings of the 29th ACM International Conference on Multimedia, 153–161. https://doi.org/10.1145/3474085.3479238.

Eidenmüller, H. (2017). The rise of robots and the law of humans. Oxford Legal Studies Research Paper No. 27/2017.

Elgammal, A. (2019). AI is blurring the definition of artist. American Scientist, 107(1), 18. https://www.americanscientist.org/article/ai-is-blurring-the-definition-of-artist.

Epic Games. (2024). MetaHuman Creator: High-fidelity digital humans made easy. Unreal Engine. https://www.unrealengine.com/en-US/meta-humans.

Epstein, Z. (2020). AI and the future of creativity. Stanford Social Innovation Review, 18(3), 40–45. https://ssir.org/articles/entry/ai_and_the_future_of_creativity.

Erdélyi, O. J., & Erdélyi, G. (2021). The AI liability puzzle and a fund-based work-around. Journal of Artificial Intelligence Research, 70, 1309-1334. https://doi.org/10.48550/arXiv.1911.08005.

Erdélyi, O. J., & Goldsmith, J. (2018). Regulating artificial intelligence: Proposal for a global solution. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 95–101. https://doi.org/10.1145/3278721.3278731.

Eshraghian, J. K. (2020). Human ownership of artificial creativity. Nature Machine Intelligence, 2(3), 157–160. https://doi.org/10.1038/s42256-020-0161-x.

Everitt, T., Lea, G., & Hutter, M. (2018). AGI safety literature review. arXiv preprint arXiv:1805.01109.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.

Fosch-Villaronga, E., & Mahler, T. (2021). Cybersecurity, safety and robots: Strengthening the link between cybersecurity and safety in the context of care robots. Computer Law & Security Review, 41, 105528. https://doi.org/10.1016/j.clsr.2021.105528.

Frith, J., & Calhoun, C. (2018). The digital city and the limits of digital citizenship: Smart cities and social equity. Information, Communication & Society, 21(1), 111-128. https://doi.org/10.1080/1369118X.2016.1209626.

Garon, J. (2022). Legal Implications of a Ubiquitous Metaverse and a Web3 Future. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4002551.

Gellers, J. C. (2021). Rights for robots: Artificial intelligence, animal and environmental law. Routledge.

Gervais, D. J. (2020). The machine as author. Iowa Law Review, 105(5), 2053–2106. https://ilr.law.uiowa.edu/print/volume-105-issue-5/the-machine-as-author/.

Gervais, D. J. (2021). The human cause. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3857844Bathaee, Y. (2018). The artificial intelligence black box and the failure of intent and causation. Harvard Journal of Law & Technology, 31(2), 889–938.

Gervais, D. J. (2021). The machine as author. IIC - International Review of Intellectual Property and Competition Law, 52(3), 341-357. https://doi.org/10.1007/s40319-021-01043-9Scherer, M. U. (2016).

Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. Harvard Journal of Law & Technology, 29(2), 353–400.

Giuffrida, I. (2019). Liability for AI decision-making: Some legal and ethical considerations. Fordham Law Review, 88(2), 439–456.

Glasser, D. (2001). Copyright in computer-generated works: Whom, if anyone, do we reward? Duke Law & Technology Review 24.

Gordon, J.-S. (2022). Are superintelligent robots entitled to human rights? Ratio, 35(3), 371-386. https://doi.org/10.1111/rati.12346

Guastavino, M. N., & Mangan, D. (2023). The metaverse matrix of labour law. Italian Labour Law e-Journal. https://doi.org/10.6092/issn.1561-8048/17138.

Gunkel, D. J. (2024). Robot rights. MIT Press.

Haney, B. S. (2018). The perils and promises of artificial general intelligence. Journal of Legislation, 45(2), 151-172. https://scholarship.law.nd.edu/jleg/vol45/iss2/1.

Hassani, H., Huang, X., & Silva, E. (2021). The human digitalisation journey: Technology first at the expense of humans? Information, 12(7), 267. https://doi.org/10.3390/info12070267.

Hedrick, A. S. (2019). I think therefore I create: Claiming copyright in the outputs of algorithms. Journal of Intellectual Property and Entertainment Law, 8(2), 324-374.

Hertzmann, A. (2018). Can computers create art? Arts, 7(2), 18. https://doi.org/10.3390/arts7020018.

Hess, C., & Ostrom, E. (Eds.). (2006). Understanding knowledge as a commons: From theory to practice. MIT Press.

Hildebrandt, M. (2020). Legal personhood for AI? In Law for Computer Scientists and Other Folk (pp. 185-206). Oxford University Press. https://doi.org/10.1093/oso/9780198860877.003.0009.

Hristov, K. (2017). Artificial intelligence and the copyright dilemma. IDEA: The Journal of the Franklin Pierce for Intellectual Property, 57(3), 431–454.

IEEE. (2019). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, first edition. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.

Izsak, K., Terrier, A., Kreutzer, S., Strähle, T., Roche, C., Moretto, M., Sorensen, S. Y., Hartung, M., Knaving, K., Johansson, M. A., Ericsson, M., & Tomchak, D. (2022). Opportunities and challenges of artificial intelligence technologies for the cultural and creative sectors. European Commission. https://doi.org/10.2759/144212.

Kamrowska-Załuska, D. (2021). Impact of AI-based tools and urban big data analytics on the design and planning of cities. Land, 10(11), 1209. https://doi.org/10.3390/land10111209.

Kim, J., & Im, I. (2023). Anthropomorphic response: Understanding interactions between humans and artificial intelligence agents. Computers in Human Behavior, 139, 107512. https://doi.org/10.1016/j.chb.2022.107512.

Kop, M. (2020). AI & Intellectual Property: Towards an articulated public domain. Texas Intellectual Property Law Journal, 28(1), 297–336.

Kostenko, O., Furashev, V., Zhuravlov, D., & Dniprov, O. (2022). Genesis of Legal Regulation Web and the Model of the Electronic Jurisdiction of the Metaverse. Bratislava Law Review. https://doi.org/10.46282/blr.2022.6.2.316.

Kostenko, O., Zhuravlov, D., Dniprov, O., & Korotiuk, O. (2023). Metaverse: Model Criminal Code. Baltic Journal of Economic Studies. https://doi.org/10.30525/2256-0742/2023-9-4-134-147.

Kurki, V. A. J. (2019). A theory of legal personhood. Oxford University Press.

Kuzminykh, A., & Rintel, S. (2020). Classification of functional attention in video meetings. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3313831.3376546.

Lehdonvirta, V. (2024). Cloud empires: How digital platforms are overtaking the state and how we can regain control. MIT Press.

Lemley, M. A., & Casey, B. (2018). Remedies for robots. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3223621.

Lessig, L. (2004). Free culture: How big media uses technology and the law to lock down culture and control creativity. Penguin Press.

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. AI and Ethics, 2(1), 127-142. https://doi.org/10.1007/s43681-022-00143-x.

Marchant, G. E., & Lindor, R. A. (2012). The coming collision between autonomous vehicles and the liability system. Santa Clara Law Review, 52(4), 1321–1340.

McStay, A. (2023). The Metaverse: Surveillant Physics, Virtual Realist Governance, and the Missing Commons. Philosophy & Technology. https://doi.org/10.1007/s13347-023-00613-y.

Militsyna, K. (2022). Legal personhood for artificial intelligence: Pro, contra, abstain? Teisė, 122(1), 148-163. https://doi.org/10.15388/teise.2022.122.10.

Miller, A. I. (2020). The artist in the machine: The world of AI-powered creativity. MIT Press.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), https://doi.org/10.1177/2053951716679679.

Mohseni, S., Wang, H., Yu, Z., Xiao, C., Wang, Z., & Yadawa, J. (2021). Taxonomy of machine learning safety: A survey and primer. arXiv. https://doi.org/10.48550/arxiv.2106.04823.

Moore, J. (2019). AI for not bad. Frontiers in Big Data, 2. https://doi.org/10.3389/fdata.2019.00032.

Nordmoen, J., Veenstra, F., Ellefsen, K. O., & Glette, K. (2021). MAP-Elites enables powerful stepping stones and diversity for modular robotics. Frontiers in Robotics and AI, 8, 639173. https://doi.org/10.3389/frobt.2021.639173.

Pagallo, U. (2018). Vital, Sophia, and Co.—The quest for the legal personhood of robots. Information, 9(9), 230. https://doi.org/10.3390/info9090230.

Qin, H.X., Wang, Y., & Hui, P. (2022). Identity, Crimes, and Law Enforcement in the Metaverse. arXiv. https://doi.org/10.48550/arxiv.2210.06134.

Rachum-Twaig, O. (2020). Whose robot is it anyway?: Liability for artificial-intelligence-based robots. University of Illinois Law Review, 2020(4), 1141–1175.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). Zero-shot text-to-image generation. arXiv preprint arXiv:2102.12092.

Russell, S. (2023). Human compatible: AI and the problem of control. Penguin.

Russell, S. J., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), 247–278. https://doi.org/10.1109/JPROC.2021.3060483.

Schafer, B., Komuves, D., Zatarain, J. M. N., & Diver, L. (2015). A fourth law of robotics?: Copyright and the law and ethics of machine co-production. Artificial Intelligence and Law, 23(3), 217-240.

Selbst, A. D. (2020). Negligence and AI's human users. Boston University Law Review, 100(4), 1315–1376.

Seymour, M., Riemer, K., & Kay, J. (2018). Actors, avatars and agents: Potentials and implications of natural face technology for the creation of realistic visual presence. Journal of the Association for Information Systems, 19(10). https://doi.org/10.17705/1jais.00515.

Shirado, H., & Christakis, N. A. (2017). Locally noisy autonomous agents improve global human coordination in network experiments. Nature, 545(7654), 370-374. https://doi.org/10.1038/nature22332.

Soares, N., & Fallenstein, B. (2017). Agent foundations for aligning machine intelligence with human interests: a technical research agenda. In The Technological Singularity. Springer.

Solaiman, S. M. (2017). Legal personality of robots, corporations, idols and chimpanzees: A quest for legitimacy. Artificial Intelligence and Law, 25(2), 155–179.

Soliman, M. M., Ahmed, E., & Darwish, A. et al. (2024). Artificial intelligence powered Metaverse: Analysis, challenges and future perspectives. Artificial Intelligence Review, 57, 36. https://doi.org/10.1007/s10462-023-10641-x.

Sotala, K., & Yampolskiy, R. V. (2015). Responses to catastrophic AGI risk: a survey. Physica Scripta, 90(1), 018001.

Srinivasan, R., & Parikh, D. (2021). Building bridges: Generative artworks to explore AI ethics. arXiv. https://doi.org/10.48550/arxiv.2106.13901Stein, A. L. (2020). Artificial intelligence and climate change. Yale Journal on Regulation, 37, 890–939.

Stix, C., & Maas, M. M. (2021). Bridging the gap: The case for an 'Incompletely Theorized Agreement' on AI policy. AI and Ethics, 1(3), 261-271.

Sun, J., Gan, W., Chao, H.-C., & Yu, P. S. (2022). Metaverse: Survey, Applications, Security, and Opportunities. arXiv. https://doi.org/10.48550/arxiv.2210.07990.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. Science, 361(6404), 751–752.

Tai, E. T. T. (2018). Liability for (semi)autonomous systems: Robots and algorithms. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3161962.

Tegmark, M. (2017). Life 3.0: Being human in the age of artificial intelligence. Allen Lane.

Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M., & Bernstein, A. (2020). Implementations in machine ethics: A survey. ACM Computing Surveys, 53(6), 1–38. https://doi.org/10.1145/3419633.

Travis, H. (2020). Intelligent entertainment: Shaping policies on the algorithmic generation and regulation of creative works. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3726755.

Turchin, A., & Denkenberger, D. (2020). Classification of global catastrophic risks connected with artificial intelligence. AI & Society, 35(1), 147–163. https://doi.org/10.1007/s00146-018-0845-5.

Turner, C., & Schneider, S. (2020). Could you merge with AI? Reflections on the singularity and radical brain enhancement. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), The Oxford handbook of ethics of AI (online edn). Oxford Academic. https://doi.org/10.1093/oxfordhb/9780190067397.013.19.

Turner, J. (2019). Robot rules: Regulating artificial intelligence. Palgrave Macmillan.

Tyagi, K. (2023). Deepfakes, copyright and personality rights: An inter-disciplinary perspective. In K. Mathis & A. Tor (Eds.), Law and economics of the digital transformation. ILEC 2023. Economic analysis of law in European legal scholarship (Vol. 15). Springer, Cham. https://doi.org/10.1007/978-3-031-25059-0_9.

van den Hoven van Genderen, R. (2018). Do we need new legal personhood in the age of robots and AI? In Perspectives in Law, Business and Innovation (pp. 185-206). Springer Nature. https://doi.org/10.1007/978-981-13-2874-9_2.

Vladeck, D. C. (2014). Machines without principals: Liability rules and artificial intelligence. Washington Law Review, 89(1), 117–150.

Wallach, W., & Marchant, G. E. (2019). Toward the agile and comprehensive international governance of AI and robotics. Proceedings of the IEEE, 107(3), 505–508.

Wolfert, P., Robinson, N., & Belpaeme, T. (2022). A review of evaluation practices of gesture generation in embodied conversational agents. IEEE Transactions on Human-Machine Systems, 52(2), 352-364. https://doi.org/10.1109/THMS.2022.3149173.

Wu, H., & Zhang, W. (2023). Digital identity, privacy security, and their legal safeguards in the Metaverse. Security and Safety. https://doi.org/10.1051/sands/2023011.

Yanisky-Ravid, S., & Velez-Hernandez, L. A. (2018). Copyrightability of artworks produced by creative robots, driven by artificial intelligence systems and the concept of originality: The formality–objective model. Minnesota Journal of Law, Science & Technology, 19(1), 1–53.

Yayman, D. (2023). Taxation in Virtual Worlds: Analysis Under United States of America and Turkish Tax Regulations. Sosyoekonomi. https://doi.org/10.17233/sosyoekonomi.2023.01.11.

Yudkowsky, E. (2016). AI alignment: Why it's hard, and where to start. Machine Intelligence Research Institute.

Zimmermann, D., Wehler, A., & Kaspar, K. (2023). Self-representation through avatars in digital environments. Current Psychology, 42, 21775-21789. https://doi.org/10.1007/s12144-022-03232-6.