Intelligent Human Systems Integration (IHSI 2023), Vol. 69, 2023, 683–691
https://doi.org/10.54941/ahfe1002891

Metavethics: Ethical, Integrity and Social Implications of the Metaverse

Matteo Zallio and P. John Clarkson

University of Cambridge, Engineering Department, Inclusive Design Group, Cambridge, CB2 1PZ, United Kingdom

ABSTRACT

This article outlines some of the major ethical and social questions raised by the birth and growth of digital, virtual environments and the metaverse. The metaverse, also known as a virtual digital space, is rapidly taking shape and forecasts from various analysts predict that this technology is going to generate a vast number of disruptive changes that can impact the privacy, safety and social sphere of human beings, giving rise to new ethical challenges and opportunities which need to be addressed. As a pioneering work, this article examines via a semantical approach the etymology and evolution of the term ethics across different domains and also the term metaverse and enquires into how ethical questions, principles and approaches can influence the design and development of a good metaverse. This article emphasises the relevance of the sociological, anthropological and philosophical dimensions of the ethical discipline in relation to digital, virtual environments and the metaverse. Metavethics, a new field of study and domain of expertise, approaches the scientific and the broader, technology-oriented communities with new questions and inspiring opportunities for the creation of digital, virtual environments that are framed within the context of acknowledging positive ethical implications for human beings. This article aims to create a foundation upon which new knowledge can be built and create a conversation around a complex and fundamental concept of the ethics of human behaviour and the metaverse.

Keywords: Metaverse, DEI, Virtual reality, Augmented reality, Digital environments, Ethics, Integrity, Inclusive design

INTRODUCTION

The metaverse has been recently described as a set of digital spaces, including interconnected immersive 3D experiences (Meta, 2022), allowing humans to have a presence in the digital world. The evident need for representation and socialisation, both in the physical and digital worlds, is a constituent part of some of the most basic human needs (Cianci and Gambrel, 2003; Maslow, 1943).

With the metaverse being a space in which people are virtually represented by avatars in a digital, virtual environment where they can connect, socialise and work, arises a series of further implications that might not be fully considered if the technological applications of the metaverse are not investigated.

Other than the nuances that characterise the development, access and use of this technology, which are of significant importance and fundamental for the creation of seamless experiences, there are a series of challenges and implications that can lead to the creation of potentially harmful digital, virtual environments (Zallio and Clarkson, 2022).

Ethical, security, safety, inclusion, accessibility and integrity implications, which are the main focus of this article, refer to the intangible nature of digital spaces which generate content that has the potential to influence people, companies, organisations and communities.

The content, both digital (e.g. an image or sound) and physical (e.g. haptic feedback, olfactory experiences) generated in these digital, virtual environments leads to far more immersive experiences than those people are used to with their current devices and apps (Srivastava, 2005). The same content may foster the creation of an idea or a change in behaviour which generates a correspondent action in both the digital and physical worlds impacting the sphere of security and safety for individuals and communities.

These implications have strong roots in the sociological, anthropological and philosophical aspects that influence the ethical impact of decisions that users as well as technologies could take. The ethical domain appears to be not deeply explored and therefore to reduce harm, increase safety and privacy, improve inclusion and accessibility of future digital, virtual environments and correlated technologies there is a need to explore the foundations of future ethical implications of the metaverse.

As a pioneering work, this article examines via a semantical approach the etymology and evolution of the term ethics across different domains and also the term metaverse and enquires how ethical questions, principles and approaches can influence the design and development of a good metaverse.

THE METAVERSE

Since the late 80s the idea of creating digital environments in which people could live a parallel life was growing niche within communities. One of the first examples dates back to 1992, the year in which Neal Stephenson wrote a science fiction novel entitled *Snow Crash* (Stephenson, 1992). In his book Stephenson describes for the first time the term metaverse as a virtual urban environment that runs around the circumference of a spherical planet.

Metaverse stands for meta 'above' and verse for 'universe' and therefore implies a meta universe or meta world in which human beings can build and behave in ways that might not be possible in the real world, giving rise to an almost infinite array of creative possibilities.

In 2003 the San Francisco-based company Linden Labs developed and released Second Life, a virtual environment in which people could deep-dive into a digital world and create a parallel virtual life with avatars and build content that other avatars could use (Linden Labs, 2003). Second Life is recognised as being one of the first attempts to build a digital, virtual environment that could exist for several years (Bobrowsky, 2022).

In more recent years large investment from the gaming industry has increased confidence and momentum in creating even more impactful and complex

digital environments that could be used with virtual reality devices such as wearable controllers and VR glasses (Warner, 2022).

The gaming and entertainment industry, including companies such as Roblox, Active Worlds, Activision Blizzard and Epic Games, began to invest heavily into developing engaging games, supported by technologies that would offer an improved experience over and above the usual keyboard and screen experience which existed previously (Microsoft, 2022).

In 2021 the well-known social media company Facebook changed its whole brand identity to Meta, to "bring together Facebook apps and technologies under one new company brand and focus on bringing the Metaverse to life by helping people connect, find communities and grow businesses" (Facebook, 2021).

Reinforcing the recent publicity surrounding the development of virtual environments that provide new experiences to potential users, a report from the Gartner research institute reported that by 2026 one quarter of people in the world will be spending at least an hour a day in the metaverse and that 30% of organisations in the world will have products and services ready for the metaverse (Gartner, 2022).

In summary, the metaverse could be described as a place of non-places, in which the digital creativity of the human brain meets the computational capacity of Machine Learning algorithms, servers and cloud services that have the potential to generate bots and other digital entities that could automate the process of creating this digital world (Dwivedi et al., 2022).

Using a somewhat unsophisticated definition it would be easy to define the metaverse as a digital copy of the real world in which users have greater freedom and can escape the reality of life. This description could constrain the potential for future innovations as well as the development of a safe, inclusive metaverse. Additionally, it may lead to future misconceptions as to what digital immersive environments are compared to physical environments, causing disorientation and misunderstanding among users and possibly result in a failure to recognise which environment they are currently inhabiting.

Undoubtedly defining the metaverse and its opportunities and challenges can assist and support with the democratisation of knowledge surrounding this topic for the wider community (Zallio and Clarkson, 2022). However it is important to highlight that in tandem with the number of opportunities that are emerging, an equal if not greater number of challenges are rising (Iqbal and Campbell, 2022).

As the definition of the metaverse and digital, virtual environments develops, the technologies, devices and software will evolve and the awareness and needs of people will grow simultaneously. These opportunities will allow people can to experience new emotions, develop new behaviours and generate new ideas unlikely to have been imagined in the physical world.

Alongside this exponential loop of creating new tools to access the metaverse, new features and immersive feedback, together with behavioural modification and a growth in user needs, an important question is raised: what is the ethical impact of all of these design-driven as well as people-driven decisions on the design of the metaverse?

Which ethical standard, policy or guideline will guide businesses to develop a good metaverse?

ETHICS

The term ethics derives from the Greek word 'ēthikós', defined as "relating to one's character", which itself comes from the root word êthos, meaning "character, moral nature" (Liddell and Scott, 1889). Ethics is a branch of philosophy that is concerned with the behaviour of individuals in society and involves systematising, defending and recommending concepts of right and wrong, along with practical reasoning such as freedom, equality, duty, obligations and choice. The Cambridge Dictionary of Philosophy states that the word 'ethics' is "commonly used interchangeably with 'morality' ... and sometimes it is used more narrowly to mean the moral principles of a particular tradition, group or individual" (Deigh, 1995).

The discipline of ethics seeks to investigate and answer questions of human morality considering concepts such as justice and crime, correct and not correct, right and wrong, good and bad (Martinez, 2020).

Ethics can refer to philosophical ethics or moral philosophy and currently there are several areas of studies of which three can be summarised as Normative ethics, Applied ethics and Meta-ethics. Normative ethics studies the pragmatic ways of determining a moral course of action (Gustafson, 2020). Applied ethics studies the obligations or permissions that a person has in a specific domain or context. Meta-ethics studies the theoretical meaning of moral propositions and how their values can be determined.

Along the original branches of ethics as mentioned above, ethics evolved and formed a strong connection with several topics related to the evolution of human beings, their society, environments and the products that they use.

Within this context, several associated disciplines such as machine ethics, the ethics of technology, computer ethics and robot ethics, have proliferated in the last century and their goal has been to explore, investigate and define ethical and social questions and implications of addressing ethical questions that are specific to technological related products or services.

Machine ethics was a term coined by Mitchell Waldrop, based on the principle that intelligent machines will possess integrated values, assumptions and purposes, whether their programmers consciously intended them to or not (Waldrop, 1987).

The ethics of technology is a subfield that addresses ethical questions specific to the age of technology. The subject has also been explored, following the work of Mario Bunge, under the term techno-ethics (Bunge, 1977). Technology ethics is the application of ethical thinking to new technologies as they continue to expand and evolve (Luppicini, 2010).

Computer ethics studies the revolutionary social and ethical consequences of information technology. The term 'computer ethics' was first introduced by Maner and refers to the field of philosophical inquiry that deals with ethical problems exacerbated, transformed or created by computer technology (Maner, 1980).

Robot ethics considers ethical problems that may occur with robots, such as whether robots pose a threat to humans in the long or short term. The field

of robot ethics has its foundations in one of the first publications that directly addressed this potential ethical concern: *Runaround*, a science fiction short story written by Isaac Asimov in 1942 which described the Three Laws of Robotics (Asimov, 1950).

All these parallel disciplines that constitute the complex and multifaceted domain of ethical studies deal with practical problems and focus on the nature of moral action and responsibility and develop knowledge-based, codified ethics which respond to a process of intentions, actions and consequences.

Similarly, with the growth of digital, virtual environments and the metaverse, new opportunities and challenges are coming to the fore in the domain of ethics, where humans as well as machines are involved in the intentions-actions-consequences loop.

Businesses, experts and researchers are in the process of uncovering several layers of complexity surrounding the study and creation of different digital worlds which will have meaningful implications on the behavioural, sociological and psychological aspects of human beings.

METAVETHICS

Several tech companies are working out the best way to design an entirely new immersive world, but when considering attitudinal, behavioural and social aspects of human diversity it is justifiable to ask how the metaverse could be truly designed by and for people (Zallio and Clarkson, 2022).

Understanding how it is possible to maximise opportunities and deliver at scale a safe, inclusive metaverse that guarantees equity and diversity with respect to ethical principles is key for the development of this technology.

By analysing via a semantical approach the etymology and evolution of the terms metaverse and ethics it is possible to identify and envisage a new term that bridges the current gap between the creation of immersive virtual environments and the sociological, anthropological and philosophical implications that could affect the safety, security, access and inclusion of people using this technology.

Metavethics arises from a need to bridge the gap between the metaverse and the opportunities this generates and the ethical needs and demands of people who will embark upon a journey where they will spend time creating and building new digital and virtual environments.

Metavethics refers specifically to the ethics of human behaviour in relation to the metaverse as it develops and grows as an advanced and pervasive immersive technology.

The main fields related to Metavethics are: computer science, artificial intelligence, philosophy, ethics, theology, biology, psychology, cognitive science, neurosciences, law, sociology, anthropology, economics and industrial design and these are constantly expanding. Experts in all these fields are heavily involved in discovery, critical thinking, analysis and definitions of the ethical challenges that new products, features and technologies embedded in the metaverse could generate for human beings.

These fields and their interconnections bring to light just how strongly the relationship between the sociological, anthropological and philosophical

aspects of this technology could have an impact in terms of ethical concerns and highlights how Metavethics can serve to provide multiple dimensions, including supporting and influencing for the better the design of digital, virtual environments and the metaverse.

Metavethics requires the combined expertise of specialists from numerous disciplines, who must explore, study and develop critical knowledge to inform the creation of the metaverse and digital, virtual environments. Additionally, one of their all-important tasks will be to create, amend and re-frame guidelines, standards, laws, policies and regulations in alignment with challenges and opportunities resulting from the scientific and technological achievements in artificial intelligence, blockchain, Web 3.0 and digital immersive technologies such as augmented reality, virtual reality and mixed reality.

Metavethics can be seen as a discipline which impacts on a sociological, anthropological, technical and philosophical level the content and design of technology, regulation, policy and governance.

One of the main questions that the discipline of Metavethics addresses is to what extent the metaverse will provide opportunities for reducing accessibility barriers and increasing inclusion, by guaranteeing a safe digital environment in which diversity and equity prevail?

The discipline of Metavethics has the potential to develop the necessary educational efforts surrounding the metaverse and act as a springboard to promote open conversations and discussion across different communities. Metavethics will assist companies and organisations in identifying new questions and test initiatives, products, technologies and services for robustness prior to them being launched in the marketplace. In turn Metavethics supports the community by affording them new opportunities and ideas in order to preserve physical and psychological safety as well as the inclusion and privacy of every individual accessing and using the metaverse.

CONCLUSION AND FUTURE SCENARIOS

Assisting the community in comprehending the metaverse as well as to understand the impact that safety, well-being, privacy and integrity have upon on people are crucial aspects for consideration in the future. Enabling people to seamlessly access the metaverse by guaranteeing social equity, diversity and opportunities for all communities to be represented is the aim of designing a metaverse that functions optimally and is inclusive for all human beings.

This foundational research into the discipline of Metavethics has highlighted a new opportunity for exploring challenges, educating communities, informing people and supporting organisations in developing a good and positive metaverse.

This new field of study and domain of expertise challenges the scientific and broader technology-oriented communities with new questions and interesting opportunities for the creation of digital spaces that are developed in such a way that acknowledge the positive ethical implications for human beings.

With Metavethics we have arrived at a very different point in history than when the Internet was first conceived. We now have greater knowledge of Inclusive Design, a better understanding of DEI practices and accessibility

standards and therefore the design community has little or no excuse to not design a metaverse and its technologies which is genuinely safe, secure, inclusive and accessible for all at the outset.

This research aims to create a foundation upon which new knowledge and expertise can be built and create a conversation around a complex and fundamental concept of the ethics of human behaviour towards the metaverse, as it develops and grows as an advanced and pervasive immersive technology.

Metavethics is not only a discipline that will inform the development of a good metaverse but will also enable a new group of experts to become metavethicists: the architects that will help organisations to build good digital, virtual environments and metaverses.

With this article we aim to stimulate conversations about a future in which the ethical dimension of humans in relation to technology is taken into account and any limitations or barriers are explored. In order to raise new questions, find solutions and illuminate fresh directions further extensive research will need to be undertaken not only by businesses but also by academic institutions, research centres and not-for-profit organisations. Metavethics creates a common ground from which experts from different disciplines can discuss and create knowledge which can be disseminated across communities, cultures and businesses.

Designing the Metaverse is an activity that has to be done by people, for people and with people (Auernhammer et al., 2022) and this will need to involve knowledge and expertise from people who designed the real world, its buildings, neighbourhoods, cities and services and lessons taken from much of what has been done before in order to guarantee representation and a culture of diversity, equity and inclusion (Zallio and Clarkson, 2021a, 2021b). In short, the designers of the Metaverse will have to learn from the past if they are to reduce the potential for mistakes and pitfalls in the future.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N°846284. Dr. Matteo Zallio conducted the research, wrote the 'Introduction', 'Metaverse', 'Ethics' and 'Metavethics', co-wrote conclusions and developed the image. Prof. P. John Clarkson co-wrote conclusions and supervised the study. All authors have read and agreed to the published version of the manuscript. Authors acknowledge the support from the Digital Education Futures Initiative from Hughes Hall – University of Cambridge.

REFERENCES

Asimov, I. (1950). 'Runaround', in Asimov, I. *I, Robot*. New York: Doubleday. p. 40.

Auernhammer, J., Zallio, M., Domingo, L., Leifer, L. (2022). 'Facets of human-centered design: The evolution of designing by, with, and for people. In C. Meinel and L. Leifer (eds.), *Design Thinking Research: Achieving Real Innovation*. Cham: Springer International Publishing, pp. 227–245. Available at: https://doi.org/10.1007/978-3-031-09297-8_12

Bobrowsky, M. (2022). Second Life Founder Returns to Take On the Metaverse. *Wall Street Journal*. Available at: https://www.wsj.com/articles/second-life-founder-returns-to-take-on-the-metaverse-11642080602

Bunge, M. (1977). 'Towards a technoethics', *The Monist,* 60(1), pp. 96–107. Available at: https://doi.org/doi:10.5840/monist197760134

Cianci, R., Gambrel, P.A. (2003). 'Maslow's hierarchy of needs: Does it apply in a collectivist culture'. *Journal of Applied Management and Entrepreneurship,* 8(2), pp. 143–161.

Deigh, J. (1995). Ethics. In R. Audi (ed.), *The Cambridge Dictionary of Philosophy*. Cambridge: Cambridge University Press.

Dwivedi, Y.K., Hughes, L., Baabdullah, A.M., Ribeiro-Navarrete, S., et al. (2022). 'Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy'. *International Journal of Information Management*, 66, 102542. Available at: https://doi.org/10.1016/j.ijinfomgt.2022.102542

Facebook. (2021). *Introducing Meta: A Social Technology Company* Available at: https://about.fb.com/news/2021/10/facebook-company-is-now-meta/

Gartner. (2022). *Gartner predicts 25% of people will spend at least one hour per day in the Metaverse by 2026*. Available at: https://www.gartner.com/en/newsroom/press-releases/2022-02-07-gartner-predicts-25-percent-of-people-will-spend-at-least-one-hour-per-day-in-the-metaverse-by-2026

Gustafson, A. B. (2020). Normative Ethics. In D. C. Poff & A. C. Michalos (Eds.), *Encyclopedia of Business and Professional Ethics* (pp. 1–5). Springer International Publishing. https://doi.org/10.1007/978-3-319-23514-1_1222-1

Iqbal, M.Z., Campbell, A.G. (2022) 'Metaverse as tech for good: Current progress and emerging opportunities'. Available at: http://dx.doi.org/10.2139/ssrn.4150689

Liddell, H.G., Scott R. (1889). *An Intermediate Greek-English Lexicon*. New York: Clarendon Press.

Linden Lab. (2003). Available at: https://www.lindenlab.com/about (Accessed: 9 February 2022).

Luppicini, R. (2010). *Technoethics and the Evolving Knowledge Society: Ethical Issues in Technological Design Research Development and Innovation*. Hershey, Pennsylvania: IGI Global. Available at: https://doi.org/10.4018/978-1-60566-952-6

Maner, W. (1980). *Starter Kit in Computer Ethics*. Hyde Park, New York: Helvetia Press and the National Information and Resource Center for Teaching Philosophy.

Martinez, V.R. (2020). 'More meaningful ethics'. *University of Chicago Law Review Online*, pp. 53–68. https://scholarship.law.duke.edu/faculty_scholarship/4196

Maslow, A.H. (1943). A theory of human motivation. *Psychological Review,* 50, pp. 370–396.

Meta. (2022). *How will Metaverse change your world? Episode 5 of let me explain has answers*. Available at: https://www.facebook.com/business/news/let-me-explain-episode-metaverse (Accessed: 9 February 2022).

Microsoft. (2022). *Microsoft to acquire Activision Blizzard to bring the joy and community of gaming to everyone, across every device*. Available at: https://news.microsoft.com/2022/01/18/microsoft-to-acquire-activision-blizzard-to-bring-the-joy-and-community-of-gaming-to-everyone-across-every-device/ (Accessed: 9 February 2022).

Srivastava, L. (2005). 'Mobile phones and the evolution of social behaviour'. *Behaviour & Information Technology*, 24(2), 111–129. Available at: https://doi.org/10.1080/01449290512331321910

Stephenson, N. (1992). *Snow Crash*. New York: Bantam Books.
Waldrop, M. (1987). 'A question of responsibility'. *AI Magazine*, 8(1), pp. 28–39. Available at: https://ojs.aaai.org//index.php/aimagazine/article/view/572
Warner, B. (2022). Why Wall Street thinks the metaverse will be worth trillions. *Fortune*. Available at: https://fortune.com/longform/wall-street-metaverse-web3-investors-roblox-meta-platforms-microsoft/
Zallio, M., Clarkson, P.J. (2021a). Inclusion, diversity, equity and accessibility in the built environment: A study of architectural design practice. *Building and Environment, 206*, 108352. Available at: https://doi.org/10.1016/j.buildenv.2021.108352
Zallio, M., Clarkson, P.J. (2021b). 'On Inclusion, Diversity, Equity, and Accessibility in civil engineering and architectural design. A review of assessment tools'. In Proceedings of the Design Society, 1, pp. 2297–2306. Available at: https://doi.org/10.1017/pds.2021.491.
Zallio, M., Clarkson, P.J. (2022). Designing the Metaverse: A study on Inclusion, Diversity, Equity, Accessibility and Safety for digital immersive environments. *Telematics and Informatics,* 101909. Available at: https://doi.org/10.1016/j.tele.2022.101909.