Dean&Francis
ISSN 2959-6157

A Comprehensive Review of Metaverse

Jiasheng Yuan

Abstract:
With the innovation of technology in modern society, the idea of Metaverse gradually emerged in the public scope. This paper provides a comprehensive review of the current Metaverse, covering a range of topics including its fundamentals, design, and hardware. A range of key concepts are identified, and the integration of artificial intelligence in Metaverse is discussed. Lastly, a real life case study and a vision into the future are provided. By giving this overview, this study provides an insight for researchers and practitioners for further development of Metaverse.

Keywords: Metaverse; Virtual Reality (VR); Artificial Intelligence (AI); Human–Computer Interaction (HCI); Immersive Technologies; Hardware Interfaces; Future Outlook

1. Introduction

1.1 Background

Metaverse has experienced rapid growth due to improvements in processing power and raised awareness. This term has been used increasingly by people, especially gamers who use virtual reality technology. Metaverse has provided more opportunities in various fields.

1.2 Objectives

The primary aim of this essay is to give an overview of key Metaverse concepts. These following details are covered.
- Overview the design of Metaverse, including the design principles of Metaverse and the relationship between transitions and human experience.
- Discuss the possible integration of artificial intelligence in Metaverse.
- Delve into a real life application of Metaverse.
- Outline the hardware devices that are used for Metaverse.
- Provide a future outlook of Metaverse.

1.3 Structure of the Paper

This paper is structured as the following.
- Section 2 covers the fundamentals of Metaverse, including its history and basic concepts.
- Section 3 delves into the design of Metaverse.
- Section 4 discusses the implementation of artificial intelligence in Metaverse.
- Section 5 gives an overview of the hardware of Metaverse.
- Section 6 investigates the future of Metaverse.
- Section 7 provides a conclusion and outlines the future directions.

ISSN 2959-6157

2. Fundamentals of Metaverse

2.1 History of Metaverse

As a relatively new notion, the concept of Metaverse started in the late 20th century.
- 1952-1990: The technology of virtual reality started to develop, with contributions by the pioneers, such as Morton Heilig and Ivan Sutherland.
- 1992: In Neil Stephenson's book Snow Crash, characters escaped from reality and entered a virtual world that is named "Metaverse".
- 2009: Digital currency, Bitcoin, was established, and Blockchain technology was invented.
- 2003: Implemented Metaverse started to appear. Pioneering platforms such as Second Life initiated opportunities for virtual socialization and online economies.
- 2016: Tim Sweeney, the CEO of Epic Games, claimed the significance of Metaverse in the future human society.
- 2020: The COVID-19 outbreak accelerated the adoption of digital lifestyle, indirectly pushing the development of Metaverse. Initiatives such as the Fortnight's virtual concerts raised the public's awareness towards Metaverse.
- 2021: Companies, such as Meta, started their Metaverse prototypes.
- By 2030: The growth of Metaverse is projected to be exponential, leading to a potential value creation of between $5 trillion and $13 trillion.

2.2 Key Concepts of Metaverse

Metaverse is a persistent virtual world that allows users to interact with the virtual content and the other users with a digital avatar. Despite its elusiveness, Metaverse's key characteristics are rather simple: immersive experience, virtual character, stable economic system, social interaction, and open creation.

2.3 Type of Metaverse

- Traditional Metaverse is an immersive experience that involves virtual reality, acting as a supplement and assistance of the real life
- Open Metaverse is a fully virtual world with no physical existence. In this realm, human consciousness is represented as a piece of code, which manipulates an avatar that contains human emotions and lives in the universe.

3. Design of Metaverse

3.1 Design Focuses

The design of Metaverse focuses on three key ideas.
- Interoperability refers to the seamlessness of the digital user experience. For example, Metaverse should allow a user to jump from one experience to another in a natural and frictionless way. Hence, transition is the key to achieve this aim.
- Immersion refers to the extent users can interact with the virtual environment. Both hardware and software technologies are needed to create virtual, auditory, and physical senses, making users feel that they are actually present in the virtual world. A natural user interaction is needed to create this feel of being there. From command line interface (CLI) to graphical user interface (GUI), and later, to natural user interface (NUI), interaction designs try to eliminate the amount of work that is needed to communicate with a device. Indeed, with less work needed, humans feel more engaged.
- Creativity emphasizes the ability for the users to create and explore in the open, shared space. Metaverse gives users freedom to express their minds to a form that can be visualised. Perhaps, with the help of artificial intelligence, more possibilities to create and explore in Metaverse can be found.

3.2 Three Design Principles

To achieve the key ideas listed above, three design principles are set.
- Narrative Composability is the freedom for users to shift from one experience to another, allowing the creation of new combinations of experience.
- Social Assortativity is the freedom for users to identify themselves, form social groups, and allow chance encounters.
- Path Discoverability is freedom for users to move spatially and discover their own paths.

3.3 Experience design

Creating an instance of Metaverse involves designing distinct experiences. There are four cornerstones to the experience design.
- Spatial tension takes awareness of the physical features of the space. Designers need to choose between open worlds, which allows users to roam around and explore, and closed worlds, which restricts user movement, based on the amount of spatial resources they have.
- Temporal tension addresses how time flows in different experiences. Time can flow contemplatively in normal interactions and reactively in gameplays. This difference in time flow can be an issue with synchronization, so designers need to choose one from the two options.
- Artifactual tension involves the divergent designs in the virtual objects. An object the Metaverse can be realistic

or fantastical. Transitioning between worlds with distinct themes can cause problems with the artifacts users bring. Hence, designers need to select an artifactual theme.

- Actor tension discusses to what extent a user can represent himself in the way he wants. The avatar design can be uniform or autonomous, raising issues for the designers when deciding.

3.4 Transition

After designing various experiences, designers need to integrate all these experiences by designing transitions. The overall transition between two experiences involves two aspects. Firstly, coherency means the similarity between the two experiences. Secondly, for it to be continuous, the transition needs to happen in the same world.

4. Metaverse and AI

4.1 Techniques of AI

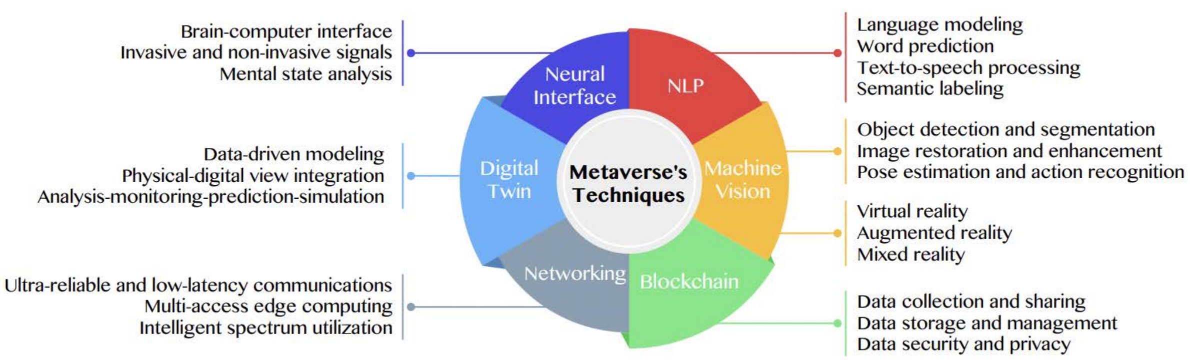
Conventional techniques use a group of data for the learning model, including supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning.

- Supervised learning uses machine learning algorithms that aims to learn the relation between inputs and outputs by mapping using labelled data.
- Unsupervised learning uses machine learning algorithms that find patterns in unlabeled data.
- Semi-supervised learning trains a model with a combination of labeled and unlabeled data.
- Reinforcement learning (RL) is a group of machine learning algorithms that makes a series of decisions in order to tackle uncertain situations.

Advanced techniques include deep learning, which is a subset of machine learning that uses a multilayer artificial neural network to simulate the complex decision making process of the human brain.

- Recurrent neural network (RNN) uses sequential data and recurrent connections to generate machine learning models.
- Convolutional neural network (CNN) uses linear algebra to find patterns in high-dimensional unstructured data.
- Self-organizing map (SOM) is an unsupervised neutral network that finds patterns in input data by reducing its dimensionality.
- An autoencoder is a type of neural network architecture that is trained to compress input data and reconstruct it to its original form.

4.2 Role of AI



Neural interface provides an interface between human brain and computer. With this interface, computers can understand brain signals and, hence, allows humans to directly give commands to the computer and receive feedback through the brain communication channel. With the neural interface, humans are able to control avatars in the Metaverse with their minds. For example, when a user wants to move his arm, his brain would send signals to his arm, and, then, his arm would respond to that signal and move. With the neural interface, the signals sent to his arm would be recognised and understood by a computer, with external monitoring devices or internally implanted devices, and a trained AI model (usually CNN) would analyse and model out the movement to allow the virtual avatar to imitate that arm movement. With this technology, humans can control the avatars in the Metaverse in a frictionless way, meaning that they do not have to process extra steps, such as commands in CLI and graphics in GUI. At the same time, users can also receive more sensational feedback using brain signals that are produced by a simulational AI model. Users can then directly involve themselves into the experience like in the real world. The simplicity of this interface in terms of communicating with the computer allows users to experience true immersiveness, which is one of the most important principles of the Metaverse. With this immersiveness, users can acquire

ISSN 2959-6157

a high level of narrative composability because they can sense the mood in the scenes and choose how to continue. A high level of interoperability can also be achieved as the user can transfer between experiences as they are actually feeling them. Furthermore, the neural interface also allows the Metaverse to be more creative. As users can directly manipulate and sense the avatar, they are able to explore more openly and do more actions that they cannot do with the older interfaces. With more actions, users can explore more things. For instance, users can be involved in an ultrareal cave exploration experience which requires them to do different movements when encountering different situations and actually feel the environment. Older input and output systems such as a mouse and a keyboard cannot provide that freedom of movement and the sensations produced by the experience. In addition, with the neutral interface, performance nuance between hardwares can be eliminated. When controlling with traditional hardware devices such as a mouse and a keyboard, users' movements and their skills largely depend on their ability to utilize the controlling device and their sensitivity. The feel of controlling also varies vastly with different devices. This difference in controlling can cause variations in user experiences and may forbid some users to be involved in some experience that requires high levels of controlling skills, lowering the overall scope of openness. With the neural interface, all users communicate in the same way, which is through the human neural network, so there would be no distinctions in the way people perceive an experience. With this openness and creativity, a high level of social assortativity can be produced, as users can choose to interact with other players in their own way, not just by how the virtual world is set to be.

Digital twin is a synchronised virtual representation of real world objects. A set of hardware devices, such as cameras, are needed to acquire a clear model of the real world scene. During this process, data-driven modeling with machine learning is used to analyse and model the scene, recognising the movements and properties of each object. Digital twin interprets behaviors of the current system and constantly monitors the performance and status of the system. It also supports the predictions of possible failures and anticipated resources using machine learning based on historical data, and it optimises the system in the virtual world before being implemented in the physical world. At the virtual end, a digital imitation would be presented, in which avatars can interact with it. Output devices, such as actuators can be implemented in the real world scene. When users use these devices in the virtual world, their operations would be synchronized to the real end. This feature is especially useful, as it allows for remote control. For instance, dangerous industrial operations can be done remotely. A set of systems would be deployed in a factory to present the work area in the Metaverse. Workers can then work using virtual reality devices, allowing them to be safe from risks. This technology provides a fully immersive system that uses real world object representations and allows for creative ideas to be tested before implemented in the real world.

Networking is the connection of computers to allow them to work interactively. It is the most essential technical aspect of Metaverse, as a high speed networking is needed for ultra-reliable and low-latency communications. Network concepts such as multi-access edge computing is used to improve the connection speed. AI technologies can be used to optimize networking. Machine learning can be used for traffic prediction and anomaly detection. Deep learning can be used for intrusion detection and traffic classification. Reinforcement learning can be used for dynamic routing and congestion control. A high-speed networking is needed to create immersive experiences in the Metaverse, as latency significantly decreases the immersiveness because it creates friction and lags for movements and perceptions. For example, latency in first-player shooting games can cause issues when battling. Latency in an online class can cause confusions, making the experience unreal. Networking is also important for interoperability. When moving between experiences, resources for the next experience need to be loaded to the local computer. This process needs to be done through the network. Hence, a fast network is needed for the transition to have a low latency and be natural. While it increases the quality of transitions, the quality of narrative composability and path discoverability can also be ensured, as users can explore and discover freely.

Natural language processing (NLP) is a part of human computer interface. It uses a variety of AI techniques including machine learning and deep learning. It is useful for word prediction. It analyses texts and generates accurate predictions of the next word. Semantic labeling is another feature of NLP. It assigns labels to words and phrases in a sentence based on their semantic characteristics. This technology also supports different features such as speech-to-text, text-to-speech, voice branding, and multi-language. More importantly, it is required for chatbots to respond to and help humans. It can also be used to generate dialogues for non-player characters in virtual worlds, enhancing the interactivity and creativity. Communications between different users can also be supported. For example, if two users use different languages, the multi-language translations feature is needed to support their communication. Hence, it also supports social assortativity, as users can freely interact and form their social groups.

Machine vision is significant in creating the virtual world and virtual reality. Machine Vision uses advanced technologies such as CNN to restore image and recognise actions. It supervises visual perception in virtual worlds, allowing XR devices (VR, AR, MR) to understand users' behaviors, detect entities, label objects, and recognize gestures. It can also be used to restore and predict some images, enhancing the freedom of creativity. For example, users can simply describe to the machine what he wants and the machine can generate it. Machine vision can also predict movements which is beneficial for experience design, especially for the ones that require many motion actions. It can increase the level of immersion in the virtual world as it can correctly label objects and allow for accurate interactions. It is also needed for a high level of social assortativity, especially in mixed reality. Other users need to be detected before any user interactions.

Blockchain is a distributed, decentralized digital ledger that has records of transactions across a network of computers. It cannot be altered without the consensus of the network. It is used to provide a secure virtual economy and information exchange. For instance, during an online transaction that uses blockchain, the transaction would be stored onto the digital ledger. Even after validations, further transactions would be executed after examining the local history by referring to the digital ledger. Machine Learning is used to detect patterns in blockchain data to find frauds. For example, it can be used to find anomalies in crypto transactions. RL is used to optimize smart contract behavior or blockchain consensus protocols. Deep learning, with more processing power, is used to process large, complex blockchain datasets. This technology can be used to predict certain risks. The security it provides can ensure a safe virtual environment for sharing personal information, which allows users to trust the Metaverse and immersively be involved in the socialization and exploration of it.

4.3 AI Fundations in Metaverse

- Natural Language Processing (NLP) uses a variety of models to understand human languages. This technology supports different features such as speech-to-text, text-to-speech, voice branding, and multi-language. More importantly, it is required for chatbots to respond to and help humans. It can also be used to generate dialogues in virtual worlds, enhancing the interactivity and creativity.
- Machine Vision uses advanced technologies such as CNN to restore image and recognise actions. It supervises visual perception in virtual worlds, allowing XR devices (VR, AR, MR) to understand users' behaviors, detect objects, and recognise gestures.
- Blockchain is a digital ledger that has records of transactions across a network of computers. It provides security for virtual economy and information exchange. AI technology is used to detect frauds to enhance cybersecurity.
- AI optimases the resource allocation for networking, ensuring low-latency.
- Digital twin is a synchronised virtual representation of real world objects. Using AI's processing power in the digital representation, real world situations can be more efficient.
- Neural Interface such as brian-machine interface (BMI) directly connects human brains to virtual interfaces. AI helps to decode the brain signals and allows interactions.
- Healthcare may use VR/AR, together with AI for analysis, monitoring, diagnosis, for remote treatment.
- AI, together with digital twins and the internet of things, cities can be managed and simulated using Metaverse.
- AI can be used to control non-player characters to provide a more immersive game experience.

4.4 Applications of AI in Metaverse

In a case of remote medical treatment, the surgeon and the patient are located in different places. The surgeon needs to conduct the operation remotely with the help of Metaverse. At the patient's location, a set of medical machines would be set up, which would often involve robots and sensors. The conditions inside the surgery room would be captured sensors and modeled by computers using AI models. This model would be uploaded onto the Metaverse. On the other side, the surgeon would have an immersive hardware set, which probably includes virtual reality devices. The Metaverse, as an agent, creates a virtual imitation of the surgery room on the surgeon's side. The surgeon can, then, start the operation. His movements would be captured by sensors and modeled by computers using AI models. This movement data then sends back to the surgery room, and the robots there operate these movements on the patient.

During this process, the neural interface can play a significant role. With the neural interface, the surgeon can control his device with actual immersiveness, meaning that the operations he does are exactly the same as how he does them in the real world. Thus, he does not need extra training to use the device, and he can operate with high precision, ensuring the patient's safety. At the same time, the neural device at the patient's side can directly monitor his mental situations and allow him to communicate with the surgeon, ensuring the quality of the surgery.

Digital twin is directly used in this case. The surgery room would be the physical world that is uploaded and monitored onto the Metaverse, and the surgeon would be

ISSN 2959-6157

the user in the Metaverse that is virtually operating. The analysis, monitoring, and prediction features of the digital twin can also help the surgeon to conduct the operation by, perhaps, fixing mistakes.

To conduct remote surgery, ultra-low latency is critical for safety, precision, and real-time control. To achieve real-time control, the latency of haptic and video transitions need to be less than or equal to 10 milliseconds. Modern systems with 5G and multi-access edge computing can reduce latency to less than 10 milliseconds, making real-time telesurgery increasingly feasible.

Natural language processing may be used to understand the surgeon or the patient's words and receive verbal commands. Using verbal commands allows them to communicate with the computer simply. Translation features may also be used when the patient and the surgeon speak different languages.

Machine vision is important as it is needed to recognise the positions, gestures, status, and motions of the patient and the surgeon. The recorded features, together with the model of the scene, are uploaded onto the Metaverse for further operations. Special cameras may be used to analyse different properties of the patient to check his real-time behavior.

Blockchain ensures security in data transmission, which can protect the surgery from malicious attacks, ensuring the patient's safety. User's data can also be secured.

5. Hardware of Metaverse

5.1 Common Hardwares

Metaverse consists of a collection of virtual spaces, a series of augmented reality (AR), virtual reality (VR), and the internet. State of the art technologies such as 5G, artificial intelligence, and wearable supporting hardware facilities also play an important role in advancing Metaverse.

5.2 Hardwares Required for AI

The technology of neural interfaces requires a range of hardware support. Different hardware is needed to capture invasive and non-invasive signals. An invasive brain computer interface collects signals that directly reflect the activities of the brain. It involves the use of brain implants that can use electrodes that can penetrate brain tissue to capture signals. Such implants include Neuralink, a neutral technology company that uses brain computer interface to translate neural signals to actions. In one of Neuralink's recent projects, an invasive brain implant allowed paralysed people to type by simply thinking. In addition, the Institute of Neuroscience of the Chinese Academy of Sciences recently established their first clinical trial of the invasive brain implant. The patient was able to use a brain computer interface to control a computer and play games such as Chess and Mario Kart Eight. Non-invasive methods involve collecting neural signals from the scalp, which provides a lower signal quality than invasive methods. This method usually uses external sensors that involve the technology of electroencephalography, measuring the electrical activity through scalp electrodes. Other sensor technologies such as functional near-infrared spectroscopy measures the blood oxygen level to determine brain activity.

The technology of digital twins relies on a combination of hardware components to collect, transmit, and process real-world data. Sensors and actuators are used to collect real-time data from the physical world and perform actions. Temperature, pressure, proximity sensors; cameras; accelerators and GPS; and actuators may be used in a digital twin system. A local computer processes data collected by the sensors, usually using machine learning models.

Networking requires reliable and fast connection, such as 5G and 6G. Multi-access edge computing is a network architecture that allows data to be computed and processed closer to the user, usually at the edge of the mobile network, such as base stations or cellular towers. Instead of sending data to a remote cloud processor, data can be processed locally, near where it is generated, reducing latency, improving response time, and saving bandwidth. This technology can be used to improve the speed of connection and network performance. Dedicated fibre connection can also be implemented to fulfil specific ultra-low latency requirements such as remote surgery. In 2001, the Lindbergh Operation (first transatlantic robotic surgery) used dedicated fibre and compression techniques to achieve a latency of 150 milliseconds.

NLP typically needs a large server to run its functionalities, so centralized processing units are needed. To capture languages from the real world, sensors are used. The most basic sensor is an audio sensor. It records analogue sound and converts it to digital signal, which can be understood by the computer. From the sounds recorded, languages can be detected and analysed. Electromyography (EMG) sensors can be used to detect muscle activity from the face to capture silent speech or subvocalization. This sensor is useful for people who have poorer abilities to speak. Cameras and motion sensors can be used to detect visual colors and vibrations to seek the spoken language. Brain computer interface sensors can directly sense the sentences spoken by analysing the neural signals.

Machine vision needs a range of hardware. Different cameras can be used to determine different properties.

3D cameras may use depth sensors. High-speed cameras may be used to capture fast-moving objects. Infrared and thermal cameras may be used to detect heat signatures for night vision and body temperature. Multispectral cameras may be used to capture non-visible wavelengths. X-ray cameras may be used to detect inner body structures. Acceleration cameras may be used to detect moving objects. A proper lighting system is needed if the chosen camera needs light for its vision.

Devices connected to the blockchain usually need a large storage and high processing power to keep updating with the digital ledger. A validation machine demands more processing power. Good CPUs and big RAMs are needed to process the requests.

6. Future of Metaverse

6.1 Possible Scenarios

If Metaverse develops as expected, four possible future scenarios may emerge.

- The Free Metaverse is open and unified, and it is constructed using open-source technologies. Users have control over their data using blockchain. It is accessible for most people and has limited or community based regulations.
- The Nerdverse is an open world with many innovations, such as neural implants, ultra-immersive VR. It is mostly used by people with more knowledge about the technology. However, due to its lack of regulations, it may not be trusted by large communities.
- Betaverses Disunited is a government manipulated universe that is small and closed, meaning that it is not connected with other Metaverses. It also involves heavy advertisements and content control.
- One Metaverse to Rule Them All is closed and unified. One large company controls and centralises everything. It has heavy censorship and surveillance.

7.2 Look into the Future

In the future, with better networking, providing ultra-low latency connection, and better blockchain technology, providing security, Metaverse can be more involved in people's lives. With improved neural interfaces, computers, with the help of AI analysing neural signals, users in the Metaverse can perceive, articulate, and interact using neural signals. For example, the virtual aspects in the virtual world would not be virtually represented to the users but sent to their brain using neural signals that model the scene. Similarly, commands would be directly received from the neural system. By then, communication between users would be simple, as they only need to exchange signals instead of translating those signals to languages. Users can directly feel and interact with the virtual world, like they are present in the world. Most of the industries would be using digital twins to have remote working and virtual simulation before operations to ensure efficiency and security.

7. Conclusion

In this paper, we provided a comprehensive review of the current and future Metaverse. Numerous fields, covering the fundamentals, the design, the hardware, the implementations of AI, and a future outlook are discussed. This paper can provide ideas for further development of Metaverse.

8. References

Copenhagen Institute for Futures Studies. (2022). *THE FUTURE OF THE METAVERSE*. https://www.massimocanducci.eu/wp-content/uploads/2023/01/Copenhagen-Institute-THE-FUTURE-OF-THE-METAVERSE.pdf

Dai, M. (2023). The Relationship between Metaverse and Design. *Highlights in Art and Design*, *3*(1), 74–76. https://doi.org/10.54097/hiaad.v3i1.9354

Ernst & Young LLP. (2023). How boards can prepare for a future in the metaverse. EY Center for Board Matters. https://www.ey.com/en_us/board-matters

Huynh-The, T., Pham, Q., Pham, X., Nguyen, T. T., Han, Z., & Kim, D. (2022). Artificial Intelligence for the Metaverse: A survey. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2202.10336

Nickerson, J. V., Seidel, S., Yepes, G., & Berente, N. (2022). Design principles for coordination in the metaverse. *Academy of Management Proceedings*, *2022*(1). https://doi.org/10.5465/ambpp.2022.15178abstract

Seidel, S., Berente, N., Nickerson, J., & Yepes, G. (2022). Designing the metaverse. *Proceedings of the . . . Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences*. https://doi.org/10.24251/hicss.2022.811