



Artificial intelligence powered Metaverse: analysis, challenges and future perspectives

Mona M. Soliman^{1,4} · Eman Ahmed^{1,4} · Ashraf Darwish^{2,4} · Aboul Ella Hassanien^{1,3,4}

Published online: 5 February 2024
© The Author(s) 2024

Abstract

The Metaverse, a virtual reality (VR) space where users can interact with each other and digital objects, is rapidly becoming a reality. As this new world evolves, Artificial Intelligence (AI) is playing an increasingly important role in shaping its development. Integrating AI with emerging technologies in the Metaverse creates new possibilities for immersive experiences that were previously impossible. This paper explores how AI is integrated with technologies such as the Internet of Things, blockchain, Natural Language Processing, virtual reality, Augmented Reality, Mixed Reality, and Extended Reality. One potential benefit of using AI in the Metaverse is the ability to create personalized experiences for individual users, based on their behavior and preferences. Another potential benefit of using AI in the Metaverse is the ability to automate repetitive tasks, freeing up time and resources for more complex and creative endeavors. However, there are also challenges associated with using AI in the Metaverse, such as ensuring user privacy and addressing issues of bias and discrimination. By examining the potential benefits and challenges of using AI in the Metaverse, including ethical considerations, we can better prepare for this exciting new era of VR. This paper presents a comprehensive survey of AI and its integration with other emerging technologies in the Metaverse, as the Metaverse continues to evolve and grow, it will be important for developers and researchers to stay up to date with the latest developments in AI and emerging technologies to fully leverage their potential.

Keywords Metaverse · Artificial intelligence · Avatar · Blockchain · Computer vision · Natural language processing · Sustainable development

✉ Mona M. Soliman
mona.solyman@fci-cu.edu.eg

¹ Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt

² Faculty of Science, Helwan University, Helwan, Egypt

³ Giza, Egypt

⁴ Scientific Research Group in Egypt, Giza, Egypt

1 Introduction

The word “Metaverse” is a fusion of “meta”, which means beyond, and “verse”, which is a shortened version of the universe. This term was first introduced by Neil Stephenson in his science fiction novel *Snow Crash* book, published in 1992 (Huynh-The et al. 2022). The book describes a virtual reality where individuals can enter and exit through their digital avatars. The concept was later expanded upon, and in October 2021, Mark Zuckerberg announced that Facebook would be rebranded as Meta to signify its shift toward building the Metaverse. This announcement brought attention to the concept of the Metaverse for sustainable development goals, which has gained interest in both academia and industry. The Metaverse is a virtual environment where people can interact in digital spaces, creating a more immersive experience. It is a persistent network of 3D virtual worlds that will eventually serve as the gateway to most online experiences and will also significantly impact the physical world. These ideas, previously confined to science fiction and video games, are poised to revolutionize various industries, such as finance, healthcare, education, and consumer products (Huynh-The et al. 2022).

Metaverse and Artificial Intelligence (AI) are two rapidly evolving technologies that are poised to revolutionize the way we live, work, and interact with each other. The Metaverse is a virtual world that is being created by merging the physical and digital worlds and is expected to provide immersive experiences and a new platform for social interaction. On the contrary, AI refers to the capability of machines to execute tasks that typically necessitate human intelligence, like discerning speech, scrutinizing images, and making decisions.

1.1 An analysis of AI in the Metaverse publications

The potential for AI in the Metaverse is vast. AI might be used to produce personalized content, more lifelike and interesting virtual settings, and even intelligent virtual assistants to assist users in navigating the Metaverse. Moreover, AI could help make the Metaverse more accessible to people with disabilities and provide enhanced security and privacy features (Huynh-The et al. 2022). Figure 1 describes the publications and citations by Web of Science in the last five years using Artificial Intelligence and Metaverse keywords in these years. It is noted that publications of AI within the Metaverse started to be active in the year 2021. The previous years (2019, 2020) had no publications that combined the two research areas.

Next is a summary and analysis of the distribution of 217 publications related to the use of AI in the Metaverse over the last 3 years. Table 1 gives a brief review of the number of publications for several research areas that involve AI within the Metaverse in the years (2021–2023). Figure 2 presents this number as a percentage value to show the most active study topics inside the Metaverse.

Based on these statistics, we can see that most of the publications come from Engineering Electrical Electronic, and Computer Science Artificial Intelligence domains, indicating the significance of these fields in exploring AI applications in the Metaverse. Computer Science Information Systems, Telecommunications, and Computer Science Cybernetics also contribute a substantial number of publications. Based on the provided statistics by Web of Science, there are some recommendations based on these statistics:

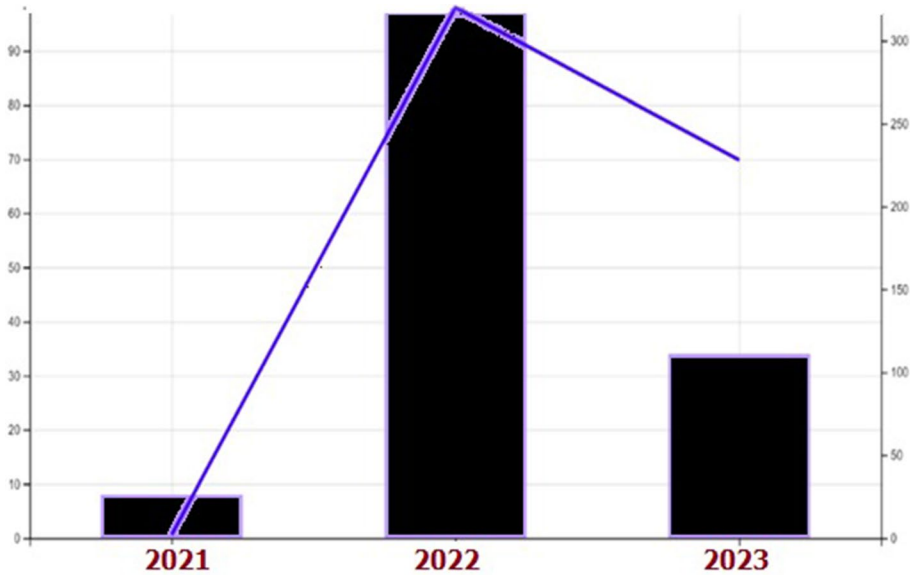


Fig. 1 The publications and citations by Web of Science in the last three years using Artificial Intelligence and Metaverse keywords from 2021 to 2023

Table 1 The number of publications for using AI in the Metaverse for the last three years (2021–2023)

Publication research area	Number of publications
Engineering electrical electronic	54
Computer science artificial intelligence	39
Computer science theory methods	18
Instruments instrumentation	16
Computer science information systems	40
Telecommunications	28
Materials science multidisciplinary	16
Applied physics	14
Computer science cybernetics	22
Computer science interdisciplinary applications	15

1. Collaboration across disciplines: Given the multidisciplinary nature of Metaverse research, researchers from diverse fields such as engineering, computer science, materials science, and physics need to collaborate. This can facilitate knowledge exchange and foster innovation in AI applications for the Metaverse.
2. Focus on AI and Metaverse integration: Given the prominence of publications in the Computer Science Artificial Intelligence field, there is a need for continued research on integrating AI technologies into the Metaverse. This includes exploring AI-driven virtual characters, intelligent environments, and personalized experiences within the Metaverse.

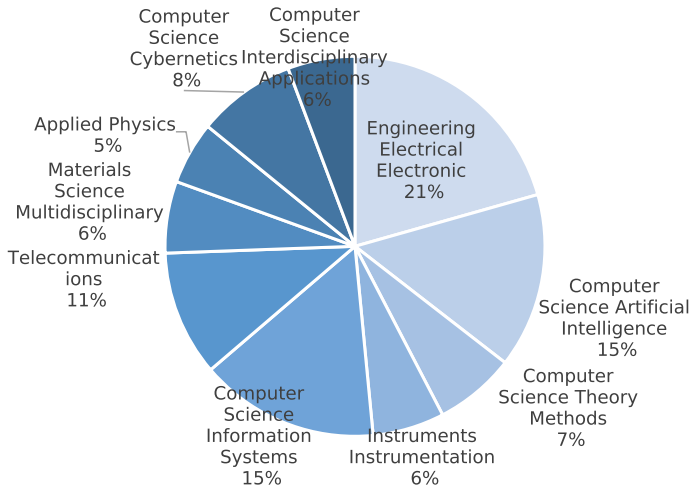


Fig. 2 The % of publications that use AI in the Metaverse for the last three years (2021–2023)

3. **Strengthening information systems and telecommunications:** The substantial number of publications in Computer Science Information Systems and Telecommunications suggests the importance of these fields in supporting the infrastructure and communication requirements of the Metaverse. Further research can be conducted to enhance network architectures, data management, and communication protocols specific to the Metaverse.
4. **Exploration of cybernetics and interdisciplinary applications:** The presence of publications in Computer Science Cybernetics and Computer Science Interdisciplinary Applications highlights the need for interdisciplinary approaches and the exploration of novel ideas in the Metaverse domain. Researchers should continue to investigate the synergies between AI, virtual reality, augmented reality, and other related fields to unlock the full potential of the Metaverse.
5. **Emphasis on practical implementations:** While academic research is valuable, it is also essential to focus on practical implementations of AI in the Metaverse. Researchers can collaborate with industry partners to explore real-world applications, address challenges, and bring innovative AI-driven experiences to users.

It is worth noting that the topic of using AI in the Metaverse is relatively recent, as indicated by the statistics gathered from the Web of Science Core Collection. With 217 publications selected over the past 3 years, it is evident that researchers and academics are beginning to recognize the importance and potential of AI in shaping the future of the Metaverse. These recommendations aim to guide researchers and practitioners in their exploration of AI applications within the Metaverse, encouraging further advancements and fostering a more immersive and interactive virtual world.

1.2 The research methodology

This subsection describes the methodology used in this paper to find the papers for our analysis on the role of AI in Metaverse applications and challenges.

We employed a systematic approach to gather relevant papers involving multiple steps. Firstly, we conducted an extensive literature review using various academic databases, including but not limited to IEEE Xplore, ACM Digital Library, ScienceDirect, Springer Nature, and Google Scholar. These platforms are widely recognized for their comprehensive coverage of scholarly articles in the field of computer science and artificial intelligence.

The initial search queries were carefully designed to capture a broad range of relevant papers related to the role of AI in Metaverse applications and challenges. The primary search terms we employed included “AI in Metaverse”, “artificial intelligence and Metaverse”, “AI applications in virtual reality”, “challenges of AI in Metaverse”, and similar variations. We ensured that our search queries encompassed relevant keywords to maximize the retrieval of pertinent literature.

Based on our initial searches, we obtained a substantial number of papers, totaling approximately 159 across the different databases. These papers were then subjected to a systematic screening process to filter out irrelevant or duplicated studies. The inclusion criteria for our analysis encompassed papers that focused on the application of AI technologies within the Metaverse context and explored the associated challenges. We excluded papers that primarily focused on AI outside the scope of the Metaverse or those that were not directly related to the central theme of our study.

To ensure the rigor and reliability of our analysis, the screening process was conducted by multiple researchers independently. Any discrepancies or disagreements in the paper selection were resolved through discussion and consensus among the research team. This approach helped us maintain objectivity and minimize potential bias in the selection process.

By following this methodology, we believe that our analysis provides a comprehensive overview of the current research landscape on the role of AI in Metaverse applications and challenges.

1.3 The scope of the proposed survey

In this paper, the fundamental role of AI in the Metaverse is analyzed and demonstrated along with Computer Vision (Shi et al. 2023), Natural language Processing (NLP) (Chen et al. 2022), and other necessary integrated technologies that must be used in the Metaverse such as the VR (Vondráček et al. 2023), AR (Qamar et al. 2023), XR (Yoo et al. 2023), MR (Fuente et al. 2022), DT (Yogesh et al. 2022) and IoT (Sun et al. 2022).

The integration of AI with the Metaverse is essential to enable people to navigate through virtual and physical worlds, and to process the massive amounts of data generated within the Metaverse (Huynh-The et al. 2022). AI needs to be built to understand the context and learn like humans, and to handle the dynamic nature of the Metaverse. While significant progress has been made, there is still a long way to go in developing a usable and commercial Metaverse system (Yogesh et al. 2022). Innovations are needed to unlock additional features and drive virtual environments closer to a perceived virtual universe. The research community is actively exploring the development of the Metaverse and the potential for new technologies to facilitate its application (Zhao et al. 2022).

There are several reasons why integrating the Metaverse with AI is important:

1. Enhanced user experience: AI might be used to produce personalized content, more lifelike and interesting virtual settings, and even intelligent virtual assistants to assist users in navigating the Metaverse.
2. Access and inclusivity: AI can help make the Metaverse more accessible and inclusive by providing features such as text-to-speech, voice recognition, and language translation, which can make the virtual world more accessible to people with disabilities or those who speak different languages.
3. Security and privacy: The Metaverse will require robust security and privacy measures, and AI can be used to enhance these measures by providing intelligent threat detection and prevention, as well as identity verification and authentication.
4. Data and insights: The Metaverse will generate vast amounts of data that can be analyzed and used to gain insights into user behavior and preferences. AI can be used to analyze this data and provide valuable insights that can be used to improve the Metaverse experience for users.
5. Automation and efficiency: AI can be used to automate certain tasks in the Metaverse, such as content creation and moderation, which can improve efficiency and reduce the workload for human moderators and creators.

The various branches of AI used in the Metaverse world are shown in Fig. 3, along with other crucially important integrated technologies. Virtual world creation, avatar generation, user interaction, and text/audio understanding are some of the different approaches utilized to transform the physical world into a visual digital world using both integrated technology and branches of AI. This survey contributes to the understanding of the integration between AI and Metaverse technologies by presenting a methodology that includes the following steps:

1. This survey is conducted on the AI techniques currently used in Metaverse technologies.
2. The role of AI techniques in computer vision, XR, and NLP is presented.

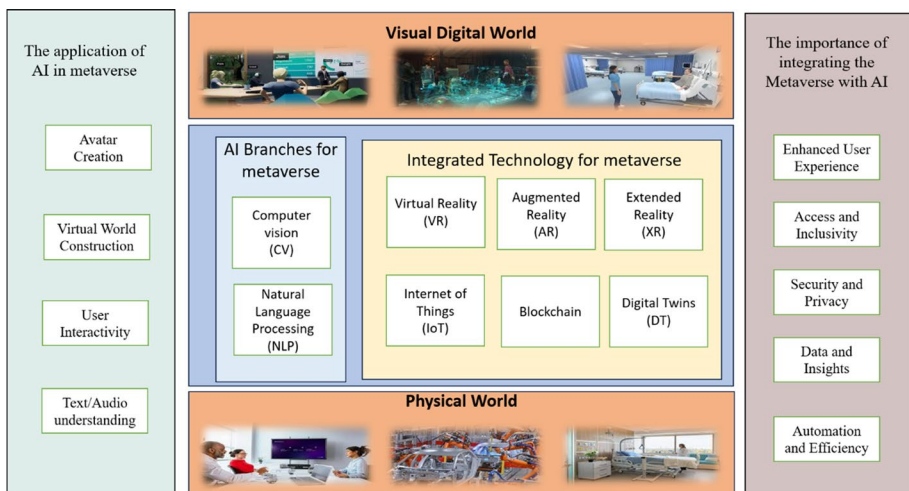


Fig. 3 The integration between AI and Metaverse technologies

3. AI techniques with integrated technologies such as blockchain, internet of IoT, VR, AR, MR, and XR are also discussed.
4. The survey links recent AI techniques with Metaverse applications to provide a comprehensive understanding of the integration between the two technologies.
5. Finally, the survey provides insights into future directions for the use of AI in the Metaverse, including potential innovations and advancements that could further enhance the integration of these technologies.

Overall, this survey offers valuable insights into the current and future state of AI in the Metaverse. The paper is organized as follows, “Sect. 2” presents the role of AI in the Metaverse applications and the state-of-the-art AI techniques used in each of them. “Sect. 3” illustrates the integrated technologies with AI in the Metaverse world. “Sect. 4” explores the usage of AI technologies in the Metaverse world. Problems and Challenges are presented in “Sect. 5”. “Sect. 6” presents the future trends. “Sect. 7” concludes this paper.

2 AI branches for Metaverse

AI is the study of technologies that enable the development of machines and computers that can imitate cognitive functions associated with human intelligence, such as sensing (Yu et al. 2023), comprehending, and responding to spoken or written language (Moradi and Shekofteh 2023), analyzing data (Salah et al. 2023), making recommendations (Pamucar et al. 2023), and more. Machine learning is a kind of artificial intelligence that enables systems and machines to automatically learn from their mistakes and improve. Instead of using explicit programming, machine learning uses algorithms to analyze vast amounts of data, learn from the results, and then make decisions based on those findings (Gokasar et al. 2023). Examples of Supervised learning include K-Nearest Neighbor (KNN), Linear Classifiers, Decision Trees, Random Forests, Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), Decision Trees, and Random Forests (Kotsiantis et al. 2007). Unsupervised learning techniques deal with unlabeled data and learn their underlying features, they aim to group them according to similarity measures. Examples of Unsupervised Learning are clustering techniques such as K-Means, hierarchical clustering, and associative rules (Hu et al. 2015).

AI plays an important role in building Metaverse worlds through being used in each of the technologies used to build Metaverse worlds and provide immersive experiences to users. Machine learning strategies can be divided into supervised learning, unsupervised learning, and reinforcement learning (Hu et al. 2015). In supervised learning, the dataset samples are labeled with the labels of their classes, the algorithms then learn the relation between the samples and their labels using the sample’s features. This way the algorithms learn how to discriminate between them by building a classifier and can recognize the label of a new sample. Reinforcement learning employs an agent that learns about its environment through applying a set of actions and receiving a reward from the environment, this enables it to estimate the value of its actions to decide the policy to deal with the surrounding environment (Kiumarsi et al. 2018). Machine learning models can also be categorized according to their objective as being Discriminative or Generative. Discriminative models aim to learn the mapping function between inputs and their output labels to be able to classify a new sample. On the other hand, Generative models tend to learn how a dataset is

generated in terms of a probabilistic model to be able to generate new similar data (David 2019).

Deep learning is an advanced branch of machine learning. Deep Neural Networks (DNNs) are ANNs with multiple hidden layers between the input and output layers. They can learn high-level features from the input data. They can be used in classification or generating data as well. Convolutional Neural Networks (CNNs) are deep learning neural networks that can learn high-level features from input images, videos, or audio datasets and classify them into different classes. They can learn high-level features from the input data. They can be used in classification or generating data as well. Convolutional Neural Networks (CNNs) are deep learning neural networks that can learn high-level features from input images, videos, or audio datasets and classify them into different classes. They have different layers: convolutional layers are in charge of extracting features from data, pooling layers are used for dimensionality reduction, and fully connected layers are responsible for making classification. Recurrent Neural Networks (RNNs) are a type of neural network that have some feedback connections connected to the layers preceding. RNNs can keep track of previous inputs thanks to the computational flow produced by the feedback connections (Lalapura et al. 2022). RNNs are considered the foundation of Long-Short Term Memory (LSTMs) and gated recurrent unit (GRU) networks. AutoEncoders (AE) and Generative Adversarial Networks (GANs) are examples of deep generative learning techniques. An autoencoder is a neural network that has been trained to map its input to the corresponding output using its two primary functions of compression and decompression (Jing and Tian 2021). The input layer of an autoencoder network has many more neurons than the hidden layer, which results in a compression of the input data into the hidden layer. As a result, the compressed version of the original input is present in this hidden layer. The output layer then performs decompression to reconstruct the input. An error function is used in the training phase to determine the difference between the input and the reconstructed output. GANs have two main neural networks, the generator, and the discriminator. The generator is trained to generate fake data and the discriminator is trained to identify the fake data from the real one. This way, GANs can learn about the dataset and produce similar samples (Jakub and Vladimir 2019).

AI has numerous applications in the Metaverse that can enhance the user experience and facilitate interactions within virtual environments. AR and VR can be improved by using AI algorithms to create more realistic and immersive environments. AI can also be used to assist with avatar creation and customization, making it easier for users to create personalized virtual representations of themselves. 3D modeling and rendering can also be enhanced by using AI, enabling faster and more efficient rendering processes. Computer vision and NLP can be used to enable more seamless interactions within virtual environments, while speech recognition can facilitate communication between users. By leveraging these AI applications, the Metaverse can become more immersive, engaging, and accessible, opening new possibilities for social interaction and collaboration within virtual spaces.

2.1 Computer vision

Computer vision is an AI subfield that allows computers to interpret and process visual inputs, such as digital images and videos, to extract relevant information and make decisions based on that information. While AI grants machines the capacity to reason, computer vision empowers them with visual perception and comprehension. Among the various categories of computer

vision are image segmentation, object detection and localization, and face detection and recognition.

- **Image segmentation:** Semantic segmentation is a task in computer vision that involves classifying an image into distinct categories based on pixel-level information (Lin et al. 2021; Lin et al. 2021). It gives the same color to all pixels belonging to a specific class. For instance, in an image having many people, all of them can have the same color, pixels of the background are assigned another color, and similarly, pixels of any object are given a certain color. Understanding the environment is considered a crucial technique, and deep learning is often employed for this purpose. A common architecture for semantic segmentation involves an encoder network followed by a decoder network (Tanzi et al. 2021). Usually, the encoder part of a semantic segmentation network uses a pre-trained classification network to reduce the size of the input image and extract important features. On the other hand, the decoder part of the network is responsible for mapping the learned distinctive features, which are at a lower resolution, to the pixel space, which is at a higher resolution, to enable dense classification. Efficient and fast segmentation of each pixel based on class information is a crucial aspect of semantic segmentation in computer vision. Deep learning-based methods, as those described in (Vázquez et al. 2020; Agarwal et al. 2021; Liang-Chieh et al. 2018), have shown considerable gains in performance on datasets for urban driving intended for autonomous driving. However, real-time semantic segmentation remains difficult to achieve with accuracy. The process of Instant Segmentation involves assigning distinct colors to different instances of the same class. Therefore, in the case of our example, each individual will be assigned a unique color.
- **Object detection and localization:** It refers to detecting different objects in an image by drawing a bounding box around them. Object localization focuses on getting the location of the detected objects whereas object recognition involves classifying the detected objects. There are two types of object detection in the Metaverse: specific instance detection (such as faces, markers, and text) and generic category detection (such as humans and cars). In XR, text detection methods have been widely researched and developed, as evidenced by previous works (Hbali et al. 2016; Nahuel et al. 2020). These methods are well-established and can be readily adapted to Metaverse applications.
- **Face Detection and Recognition:** Face detection refers to identifying faces by drawing bound boxes surrounding them whereas face recognition represents recognizing the human face. Recent years have witnessed extensive research in face detection, with robust methods demonstrated in various recognition scenarios in XR applications, as shown in (Xueshi et al. 2020; Tanja et al. 2020; Amin et al. 2020; Bernardo et al. 2019; Jan and Jonatan 2018). In the Metaverse, where people are represented by Avatars and can converse with one another, face detection and identification algorithms must distinguish between real and fake faces. Face detection in the Metaverse is further complicated by the difficulties of occlusion, sudden posture changes, and variations in illumination. Detection of generic categories has also been extensively studied, with a significant focus on using deep learning methods to detect multiple classes.

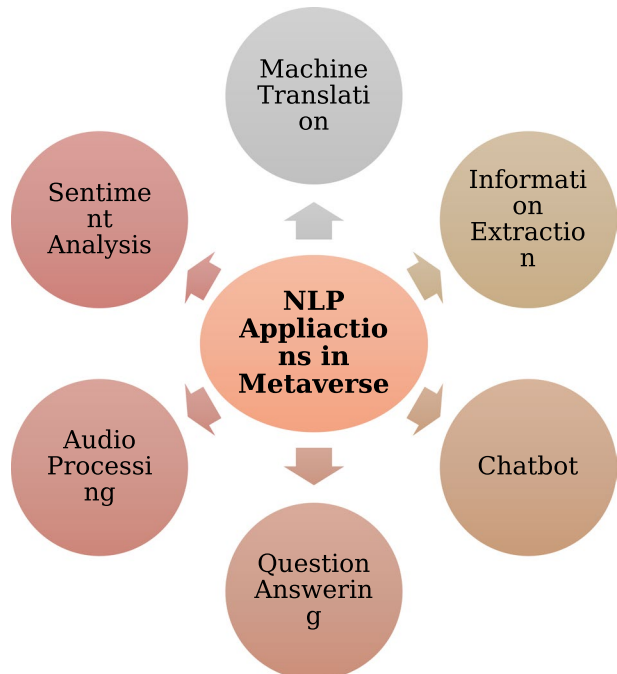
2.2 Natural language processing

NLP plays an important role in enabling more natural and intuitive interactions between humans and virtual environments in the Metaverse. With NLP, virtual assistants and chatbots can understand and respond to user input in a more human-like way, enabling more

engaging and immersive experiences. NLP can also be used to enable virtual environments to process and analyze large amounts of textual data, such as user reviews or social media posts, to better understand user preferences and behavior. In addition, NLP can be used to support multilingual interactions, allowing users from different linguistic backgrounds to communicate and collaborate within virtual environments. The importance of NLP in facilitating natural language interactions and analysing textual data is projected to increase as the Metaverse continues to develop and grow. Some applications of NLP in Metaverse are presented in Fig. 4. The Metaverse makes it easier to keep track of daily activities in work and social life, which could help NLP, a field of study that receives a lot of support from AI. As shown in Fig. 4, NLP can be used in the Metaverse to provide many tasks like machine translation, information extraction, chatbot, Question Answering, audio Processing, and sentiment Analysis. In a hybrid society, learning how to connect with people from many backgrounds and eras could be a useful application of AI (Cheng et al. 2022). In the following, we explain the key NLP applications used in the Metaverse, these include machine translation, text-to-image and image-to-text, multimodal data processing, and large language models. After that, how NLP is applied to the Metaverse is illustrated.

- **Machine Translation:** When people from different countries communicate with one another, it is essential to support many languages. A person cannot possibly learn every language spoken on the globe. The key role of AI today is to carry out automatic machine translation across different languages. The system automatically translates the user's content into any other target language while he speaks and writes in his native tongue. A problem is when there is a shortage of data for a certain language. In an attempt to cover all languages, Meta AI introduced the No Language Left Behind (NLLB) (Costa-jussà, et al. 2022) model. To effectively train AI models, it focuses on

Fig. 4. NLP applications in the Metaverse world



transforming low-resource languages into high-resource ones. First, they developed the Flores-200 many-to-many multilingual dataset, which enables testing translation quality. Then, they employed a distillation-based sentence encoding method called LASER3 to mine web data to produce parallel datasets for low-resource languages. Multilingual Mixtures-of-Experts models were trained to high performance levels using a combination of mined data and a specified set of human-translated seed data. The final model, which was trained on twice as many languages as previous models, performed 40% better than the prior state-of-the-art on the Flores-101 dataset. They developed a library of toxic words for all 200 languages using a combination of automatic and human evaluations to identify and prevent potentially damaging translations that are hallucinated by the translation algorithms. To aid in future studies, they created open-source benchmark data, scripts, and models (NLLB Team 2022). Another strategy to improve machine translation for low-resource languages is to transfer the knowledge obtained by training a model with a high-resource language pair to another model with a low-resource language pair. The use of transfer learning for neural machine translation of the English-Khasi language pair has been discussed in (Hujon et al. 2023), where the experiments and enhancements are reported.

- **Text-to-Image and Image-to-Text:** Text-to-Image enables us to create realistic images from text descriptions. To achieve this, two main steps must be taken. The first step is understanding the text and the second one is creating the images. The first step involves language modeling and extracting meaningful features to help in generating the target images. The second step should produce semantically consistent and visually realistic images. GANs are widely used for text-to-image synthesis. GAN-INT-CLS (Reed et al. 2016) was proposed to create images based on the encoded text description. In (Talasila et al. 2022), Talasila et al. proposed to use of an optimized GAN that produces images using textual features extracted using BI-LSTM. They used the Dragon Customized Whale Optimization (DC-WO) model to choose the optimal weights for GANs. Self-supervised learning is a type of machine learning that utilizes unlabeled data to enhance the learning capacity of a model by predicting other samples, eliminating the need for additional labeled data. This method was used in (Tan et al. 2022) to create more realistic images. Self-supervised learning enhances the performance of the discriminator by providing it with more images. This helps in stimulating the generator to output more diversified images. DALLE-2 was introduced by OpenAI (Ramesh et al. 2022; OpenAI 2022) as a text-to-image generative model that comprises two primary components: a prior that generates a CLIP image embedding based on a text caption and a decoder that creates an image conditioned on the image embedding. The image-to-text is the opposite of text-to-image, and it involves generating descriptive text from input images. Image captioning is a popular application of this technique. In the basic approach, visual features are extracted from the image or video, and these features are used as input to a recurrent network that generates words sequentially. Another approach is to use generative models like Variational Auto Encoders (VAEs) or GANs for generative image captioning. Further information on this topic can be found in a recent survey (Żelaszczyk and Mańdziuk 2023).
- **Multimodal Data Processing:** In the hopes that its so-called multi-modal systems will power its AR and Metaverse products, Meta has built a single AI model that can analyze audio, images, and text. The data2vec model (Baevski et al. 2022) is capable of carrying out a variety of tasks. It can identify speech from an audio sample. It can categorize items if given an image. When presented with text, it may analyze the tone and emotions of the writing or check the grammar. Data2vec is learned on three dif-

ferent modalities, unlike other AI systems, which are normally taught on just one form of input. Nevertheless, it continues to process sounds, pictures, and text individually. Data2vec enables the prediction of the representation of the/ input data, thus, a single algorithm can work with completely different types of input. Transformer, which is deep learning architectures is used also for multimodal data processing; for instance, AOBERT (Kim and Park 2023) is a multimodal single-stream transformer that predicts sentiment by using different modalities.

- Large language models: large language models, also known as LLMs, have been a game-changer in NLP as they have significantly increased the capabilities of text-based systems. These models contain a vast number of parameters, are typically several gigabytes in size, and are trained on massive amounts of text data. One of the most impressive LLMs currently available is GPT-3 (Leippold 2022), which is the third version of OpenAI's GPT language model and has a staggering 175 billion machine-learning parameters. GPT-3 can be adjusted for a variety of NLP tasks, including language translation, question-answering, and text summarization. ChatGPT (OpenAI 2023) is a variant of GPT-3 that was introduced in November 2022. It generates conversational responses to asked written questions. The training of the model involves a combination of reinforcement learning algorithms and human input, using more than 150 billion parameters. ChatGPT is focused on providing a human-like chat experience, which makes it an intelligent virtual assistant that can be used in several applications including the Metaverse (Ben 2023). Other LLMs include Megatron-Turing NLG (Developers 2023) produced by NVIDIA and Pathways Language Model (PaLM) introduced by Google (Chowdhery 2022).

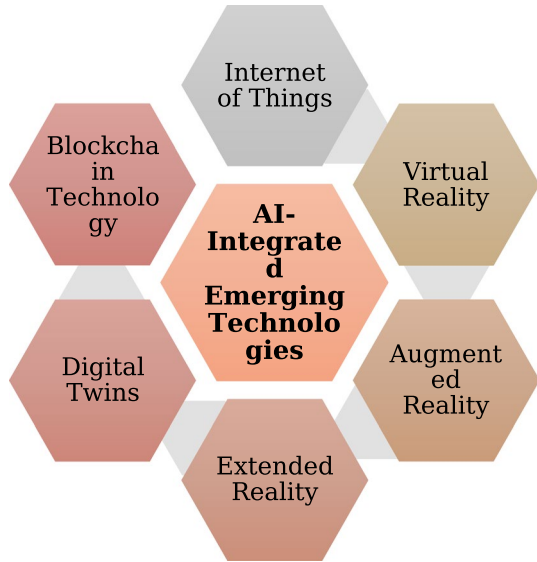
3 AI-integrated technologies for Metaverse applications

As the Metaverse continues to develop and evolve, AI-integrated technologies are expected to play an increasingly important role in shaping the way we interact with virtual environments. VR, AR, and MR technologies all offer unique opportunities for immersive and engaging experiences in the Metaverse, and AI can be used to enhance these experiences through personalized content recommendations, adaptive interfaces, and more natural and intuitive interactions (Huynh-The et al. 2022; Yogesh et al. 2022; Zhao et al. 2022). Additionally, the IoT can be used to connect virtual and physical environments, enabling seamless transitions between the two and providing more data for AI algorithms to analyze and learn from. Blockchain technology, on the other hand, can be used to create secure and decentralized virtual economies within the Metaverse, enabling users to buy, sell, and trade virtual goods and services (Yiu 2021). Figure 5 depicts the integrated emerging technologies with AI for Metaverse applications.

3.1 Virtual reality emerging technology

The integration of AI with VR has the potential to transform the way we interact with the Metaverse by enabling more natural and intuitive experiences. By using AI algorithms to analyze and learn from user behavior and preferences within virtual environments, developers can create more personalized and adaptive VR experiences. For example, AI can be used to track user gaze and gestures within VR, enabling more natural and intuitive interactions with virtual objects and environments. Additionally,

Fig. 5 AI-integrated with emerging technologies for Metaverse applications



AI can be used to create more realistic and immersive environments within VR, such as simulating changes in lighting or sound based on the user's behavior and preferences. By integrating AI with VR, developers can create more intelligent and responsive Metaverse applications that offer a range of benefits for users and businesses alike, including improved user experiences, increased efficiency, and enhanced security.

A 3D replicated environment of the actual world that enables the user to fully immerse themselves in it is what is known as virtual reality (VR). VR comprises four primary elements: virtual world, immersion, sensory feedback, and interactivity. A virtual world has basic components, which are a 3D scene, Avatars, and 3D virtual objects. A scene refers to whatever a user can see in the virtual environment from his/ her position and point of view. It may reflect a real-world place such as a virtual campus or a virtual museum. 3D virtual objects can mimic physical ones. It should have a similar appearance and texture based on its material.

VR uses graphics techniques to build its components. It needs rendering and view synthesis. The process of generating a visual representation of a 3D model is known as rendering. The model incorporates various features, including shading, textures, lighting, shadows, and viewpoints. The rendering engine's role is to process these features to produce an image that is as realistic as possible, as stated in (Datagen 2023). The rendering phase is crucial since it allows virtual objects to be presented with actual features, making them appear tangible to us. There are three prevalent rendering algorithms: ray casting, which uses basic optical principles of reflection to create an image from a specific viewpoint; rasterization, which projects objects geometrically based on model data without optical effects; and ray tracing, which uses Monte Carlo techniques to generate a realistic image in significantly less time. NVIDIA GPUs leverage ray tracing to enhance rendering performance. Color information, which is necessary in many applications, cannot be obtained using laser scanning. A hybrid 3D reconstruction based on scan data and images is employed to colourize the point cloud data. Ma et al. present a differential framework example for freestyle material capturing (Ma et al. 2021). This technique improves sampling efficiency and enhances the precision of material

properties, which in turn enables the creation of more realistic virtual objects in the Metaverse.

Additionally, modelling and rendering can be carried out with the help of a variety of software tools, which can be divided into three categories. First off, engineers and designers can make models by entering real-world data, such as the materials and weight, using parametric 3D modelling or CAD (Computer Aided Design) tools like AutoCAD and SketchUp. Additionally, polygon models made using 3D Max, Maya, or Blender are more concept-focused than measurement-focused. Thirdly, compared to polygon modelling, some digital sculpture modelling software demands a higher level of artistic talent. To create virtual worlds, these modelling tools can be integrated with engines like Unity3D or Engine. Computer graphics software is also used for design-based animation to create the illusion of objects moving in 3D space.

Volume rendering is a method used to generate a 2D representation of a 3D discretely sampled dataset. When a camera is positioned, a volume rendering algorithm obtains the RGB α (Red, Green, Blue, and Alpha channel) values of each voxel in the region where the camera's rays intersect. The RGB α color is then transformed into an RGB color and stored in the corresponding pixel of the 2D image. Every pixel is subjected to the same procedure until the full 2D image has been rendered. In contrast to volume rendering, view synthesis involves generating a 3D view using a sequence of 2D images. This can be accomplished by capturing several images of the object from different viewpoints, constructing a hemispheric layout of the object, and positioning each image in the appropriate location around the object. Given a collection of photos that depict various angles of an item, a view synthesis function tries to forecast the depth. View synthesis can be used to construct Avatars, which are virtual representations of characters from the Metaverse, such as the user and other Non-Player Characters (NPCs).

Mildenhall et al. (Ben et al. 2020) created Neural Radiance Fields (NeRF) to generate 3D views of intricate scenes with limited input views by optimizing a continuous volumetric scene function. NeRF is a fully connected Deep Neural Network (DNN) that is not convolutional. A single continuous 5D coordinate vector, comprising spatial location (x, y, z) and viewing direction (θ, ϕ) , is used to output the volume density and view-dependent emitted radiance at the given spatial location. It is trained to replicate input views of a scene using a rendering loss function. It functions by interpolating between input photos of a scene to create a single rendered scene. NeRF is a slow and computationally demanding approach that can only be used with static situations. NSVF, or Neural Sparse Voxel Fields, is a novel neural scene representation for quick and excellent free viewpoint rendering that was introduced in (Lingjie et al. 2020). To model local attributes in each cell, NSVF defines a set of implicit fields that are voxel-bounded and arranged in a sparse voxel octree. With the use of a differentiable ray-marching process from a collection of posed RGB photos, it is possible to learn the underlying voxel structures. Voxels without scene-related content are skipped. Compared to NeRF, it is 10 times faster. NeRF rendered each pixel using a single ray, which frequently resulted in blurring or aliasing at different resolutions. Mip-NeRF, described in (Jonathan et al. 2021), renders each pixel using a conical frustum rather than a ray, which lessens aliasing, enables the display of fine details in an image, and lowers error rates by between 17 and 60%. Additionally, the model outperforms NeRF regarding speed and error rates. NeRF's issue of slow rendering is addressed by KiloNeRF (Christian et al. 2021). Instead of using a single large DNN that needs to be queried repeatedly, KiloNeRF distributes the workload among thousands of little DNNs. To enhance performance and reduce storage requirements while achieving superior visual quality, NeRF employs small DNNs that represent distinct segments of the

scene. Plenoptic voxels (Plenoxels) (Alex et al. 2021) replace the central DNN in NeRF with a sparse 3D grid. The voxels surrounding each query point are used to interpolate and calculate its value. As a result, rendering new 2D views is accomplished without the need for a neural network, considerably reducing complexity and computational demands. Plenoxels are two orders of magnitude faster than NeRF and offer a similar visual quality. Inverse rendering is applied using the original NeRF in Instant-NeRF (Jonathan 2022; Isha 2022), which was made available by NVIDIA.

3.2 Augmented reality emerging technology

The integration of AI with AR has the potential to revolutionize the way we interact with the Metaverse by enabling more natural and intuitive experiences. By using AI algorithms to analyze and learn from user behavior and preferences within AR environments, developers can create more personalized and adaptive AR experiences. For example, AI can be used to track user gaze and gestures within AR, enabling more natural and intuitive interactions with virtual objects and environments. Additionally, Real-time object tracking and identification made possible by AI make it possible for virtual items to fluidly interact with the user's physical environment, resulting in more lifelike and immersive AR settings. By integrating AI with AR, developers can create more intelligent and responsive Metaverse applications that offer a range of benefits for users and businesses alike, including improved user experiences, increased efficiency, and enhanced security.

It gives users a realistic holographic, video stream, and graphic experience. When the AR-enabled game Pokémon Go first debuted in 2016, it allowed players to interact with Pokémon characters by superimposing them onto their environment using their smartphone cameras. AR technology can be delivered through smartphones, screens, glasses, head mount displays, and AR software. There are two main categories of AR, which include marker-based and markerless AR. Marker-based AR applications utilize image recognition to detect a marker, an object that triggers the AR application. On the other hand, markerless AR applications do not require markers and allow users to choose the location where they want to display digital content. To gather information about the environment, markerless AR applications rely on the device's camera, GPS, compass, and accelerometer. These applications can be categorized as Location-based AR, which offers data based on the user's location (Yassir and Salah-ddine 2020; Nextech 2022).

The Metaverse, which can be viewed as a physical embodiment of the internet, can be integrated with the real world using augmented reality (AR). Object detection is the primary technology utilized in AR, allowing computer technology to identify and locate objects of interest in images or videos (Zhang et al. 2021). There are two types of object detection techniques which are 2D object detection and 3D object detection. The former aims at drawing a 2D bounding box surrounding the object whereas the latter draws a 3D bounding box and takes into consideration the pose and exact position in the 3D space. 3D object detection can be done in many ways including estimating the 3D bounding box from 2D images, from point clouds generated by 3D laser scanners such as LiDar, and from learning features from the point clouds.

A unified architecture called PointNet (Qi et al. 2016) learns both global and local point properties. To detect 3D objects, PointNet is used in (Qi et al. 2017) to learn the point cloud features straight from the raw point cloud. Pixor (Yang et al. 2019) used a 2D CNN for feature extraction and projected a point cloud onto a map for a bird's-eye perspective. VoxelNet (Zhou and Tuzel 2018) is a network that uses 3D convolution

to obtain high-level features after performing feature learning from raw point clouds. Then, to perform classification and bounding box regression, a region proposal network is incorporated. VoxelNet training and inference are accelerated using sparse convolution (Yan et al. 2018).

Convolutional Attention Mechanism (CAM) is suggested by (Zhang et al. 2021) for 3D object detection. The first and last layers of the CNN convolution are enhanced by CAM using a channel attention mechanism and a spatial attention mechanism, respectively. Reducing the effects of illumination and occlusion on the RGB images used to create the 3D object detection, improves the feature extraction process. Polar-Mix (Xiao et al. 2022) is a data augmentation technique for point clouds generated by LiDAR to enhance the performance of deep learning techniques.

3.3 Mixed reality emerging technology

Through the creation of more immersive and interesting experiences, the combination of AI and MR has the potential to completely change how we engage with the Metaverse. Developers may produce more individualised and adaptive MR experiences by analysing and learning from user behaviour and preferences within MR settings using AI algorithms. For example, AI can be used to track user gaze and gestures within MR, enabling more natural and intuitive interactions with virtual objects and environments. Additionally, Real-time object detection and tracking made possible by AI makes it possible for virtual items to interact with the user's actual environment in more lifelike and dynamic ways in MR environments. By integrating AI with MR, developers can create more intelligent and responsive Metaverse applications that offer a range of benefits for users and businesses alike, including improved user experiences, increased efficiency, and enhanced security (Ford et al. 2023 Mar). Furthermore, the combination of AI and MR technology can facilitate the development of new use cases and applications for the Metaverse, such as immersive training and education experiences, collaborative workspaces, and gaming environments.

VR systems immerse users in completely virtual environments whereas AR overlays digital content on the top of the real world without taking into account its dynamic structure. On the other hand, MR systems constantly collect new information about the status of the environment. Accordingly, MR systems involve adding environment-aware 2D/3D content into the real world. It can present a transition experience between VR and AR as virtual objects can be superimposed on the real-time dynamic real world.

Object detection plays a crucial role in achieving the Metaverse and is widely used in XR. Face detection is a common task in VR, whereas text recognition is prevalent in AR. Advanced AR applications may require object recognition to link a 3D model to the physical world, which necessitates accurate object detection and class recognition algorithms. By linking a 3D virtual object to a physical object, users can manipulate and move it, creating a more immersive 3D environment in the Metaverse. Initial attempts at semantic segmentation relied on feature tracking methods such as SIFT, which segment pixels based on handcrafted features and utilize classification techniques like SVM. Even though they are used in both VR (Jenny et al. 2016) and AR (Peer et al. 2019), these techniques have poor segmentation performance. The potential of CNNs for semantic segmentation has recently been studied (Huanle et al. 2020).

3.4 Extended reality emerging technology

The integration of AI with XR has the potential to revolutionize the way we interact with the Metaverse by enabling more natural and intuitive experiences (Yogesh et al. 2022). By using AI algorithms to analyze and learn from user behavior and preferences within XR environments, developers can create more personalized and adaptive XR experiences. For example, AI can be used to track user gaze and gestures within XR, enabling more natural and intuitive interactions with virtual objects and environments. Developers can construct more responsive and intelligent Metaverse applications by combining AI and XR, which has a number of advantages for both users and enterprises, including better user experiences, increased productivity, and improved security (Dirk et al. 2021). Additionally, AI and XR can offer a framework for creating fresh and inventive Metaverse use cases, such as virtual healthcare, remote offices, and immersive entertainment.

The goal of XR is to give the user a seamless transition between fully virtual and fully real worlds. As a result, it connects the two worlds. It includes MR, AR, and VR under its wing (Venkatesan et al. 2021). Metaverse visual world construction embraces these different technologies and their corresponding tasks.

In VR/AR/XR systems, optical see-through or video see-through displays are the principal means of gathering visual data. Following processing, users are presented with this visual data via head-mounted displays or cellphones. In this context, computer vision technology plays a crucial role in the analysis, interpretation, and processing of visual information in the form of digital images and videos to make informed decisions and take appropriate actions.

We anticipate that human users will be represented in the Metaverse by avatars and followed by computer vision algorithms. Visual world Creation tasks which are critical to the success of every Metaverse application can be divided into three main parts: scene generation and recognition, non-player character (NPC), and Player character (Avatar) construction.

By integrating these technologies with AI, developers, and researchers can create more immersive, secure, and personalized Metaverse applications that offer a range of benefits for users and businesses alike.

3.5 Internet of things

The integration of AI with the IoT has the potential to revolutionize the way we interact with the Metaverse by enabling more seamless and intuitive experiences. By connecting virtual and physical environments through IoT devices, AI algorithms can analyze and learn from vast amounts of data in real-time, enabling more personalized and adaptive Metaverse applications. For example, IoT devices can be used to track user behavior and preferences within virtual environments, allowing AI algorithms to recommend personalized content and services based on this data. IoT devices can also be utilized to build more lifelike and immersive Metaverse environments, such as ones that simulate changes in temperature or lighting based on the user's actual environment. Developers may build more responsive and smarter Metaverse applications by fusing AI and IoT, which has a number of advantages for both users and enterprises, including better user experiences, increased efficiency, and improved security.

The term IoT refers to a broad idea that encompasses a wide range of sensors, actuators, data storage, and processing capabilities. As a result, any IoT-capable gadget may sense its surroundings, communicate, store, and analyze the data that it gathers, and take suitable action. All that occurs in the final stage of acting in the agreement is determined by the processing stage. An IoT service's actual level of intelligence can be determined by the amount of processing or action it is capable of. A non-smart IoT system will be unable to adapt to changing data and will have limited usefulness. Nevertheless, a more advanced IoT system will incorporate AI and may be able to achieve the actual objective of automation and adaptation. (Ghosh et al. 2018). The study paper (Wu et al. 2020) discusses the security of IoT and how AI may improve it. New security technologies are urgently needed due to the special characteristics of IoT security and the limitations of current solutions. As a new technology path, AI provides a wide range of applications. ML is a subject of research in the field of AI. Its theory and techniques have been applied to a variety of engineering difficulties to solve challenging issues. IoT security uses two classes of ML algorithms: transactional and decisional. Research (Kuzlu et al. 2022) explored how AI will impact the IoT revolution. Several intelligent tasks, like speech recognition, language interpretation, dynamics, and others, are anticipated to be performed by simulated intelligence without the involvement of a human. On the other hand, the IoT links a network of networked objects that communicate with one another across a network. IoT technology has entered our daily lives to enhance our level of comfort. These web-based gadgets provide unfathomable amounts of data that may be used to understand customer behaviors, styles, unique information, etc. They cannot be ignored. Many projects don't know how to store and cycle such enormous amounts of data, though. This is impeding the IoT's ability to grow and expand. In this case, the ability of man-made intelligence to sort through the avalanche of data produced by IoT devices may be of great use. It enables you to evaluate the information and base your conclusions on what you discover.

A promising strategy for creating multimodal interaction systems that provide engaging user experiences, especially for non-expert users, is seen to be the integration of IoT with AR/VR/MR. This is because it enables interaction systems to combine immersive AR content with the agent's real-world context. However, to fully realize the potential of the Metaverse, advanced hardware technologies such as VR and AR devices are necessary. While it is still possible to interact with the Metaverse using smartphones and laptops, these devices will eventually become outdated. In order to respond effectively in 3D virtual environments, a human avatar, for example, must be able to recognize the movements of other avatars or objects. Human avatars must also be able to understand the psychological and emotional characteristics of others as well as the 3D virtual world in the real world.

In order to enable wireless and seamlessly connected immersive digital experiences, the Metaverse (Kanter 2021) can employ the IoT to transform real-time IoT data from the physical world into a digital reality there. The IoT can improve how users physically engage with the virtual world created by AR/VR. Take the health awareness app as an example. Medical IoT devices that can be mounted to the user's body, or a body suit equipped with sensors can instrument the user's state, such as health issues that may cause a response in the virtual world (Pereira et al. 2021). The use of IoT sensors to measure a user's body movement in virtual fitting rooms can also improve e-commerce experiences. The data from smartphone images or smart scales, for example, might be used to update the user's personal bodily information. This goes beyond the restrictions of typical online purchasing by letting the Metaverse users immerse themselves in a virtual version of the store. The latest Tactile Internet, which creates a network or network of networks for humans or machines to remotely access or manipulate real or virtual things in real-time, can also

make use of IoT data (Sodhro et al. 2018; Promwongsa et al. 2020). In addition to facilitating data transmission between the digital and physical worlds, IoT data can give AR/VR applications context and situational awareness of physical objects (Aijaz and Sooriyabandara 2018; Lu et al. 2021). An augmented reality device, for instance, might react to the user's finger motions or activate a cyber-physical feature in reaction to a real-world event. The interaction between the real world and the virtual one made possible by the Internet of Things (IoT) results in the creation of a digital twin, or a digital representation of the actual state of a specific physical thing (Minerva et al. 2020). The Metaverse wants to make sure the reflection is as near to the actual physical condition as feasible to create a useful digital twin. Digital twins become one of the core uses in the Metaverse as a result of this distinctive trait. To make a group meeting effective in a professional setting, digital twins can be created using the Tactile Internet and Haptic Codecs (IEEE P1918.1.1) (Steinbach et al. 2018). Users can communicate with one another while using or exhibiting a replica of the hardware or software prototype. Digital twins aid engineers in operating 3D simulations of complicated systems during technical training programs (Stojanovic and Milenovic 2018). The digital twins offer a simulated environment that is connected to a real-world physical workplace where upkeep can be performed remotely. Urban planning and construction can virtualize a real-world city using digital twins, enabling locals or commercial players to use visual representation to carry out a development plan and identify prospective future urban projects (Farsi et al. 2020).

The Internet of Things is made possible by the use of wireless sensors, which also give AI the essential data it needs to build a digital environment that is useful to people. When wireless sensors are employed extensively, powering them has become a pressing issue. The most popular power source up to this point has been a battery, which can run the sensor for a predetermined amount of time and store a specific amount of electric energy. The battery must be recharged or replaced once the electrical energy it contains is expended, which is an unmanageable load for the growing number of sensors. An efficient alternative is to use self-powered sensors, which can draw energy from their environment (such as light, heat, and mechanical vibration) to power themselves (Li et al. 2023). Research work in (Gokasar et al. 2023) and (Muhammet et al. 2023) investigate the capabilities of the Metaverse to traffic safety by employing self-powered sensors. In (Gokasar et al. 2023) Metaverse self-powered sensors can capture uninterrupted data that allow for activities such as the management of the traffic network, the optimization of transportation facilities, and the management of urban and intercity journeys to be performed. (Gokasar et al. 2023) uses a case study based on a densely populated metropolis with an extensive education system. The findings of the study in (Muhammet et al. 2023) show that public transportation is the most appropriate area for implementing the Metaverse into traffic safety because of its practical opportunities and broad usage area.

The coordination of various elements, such as objects, avatars, and their interactions, is essential in virtual shared spaces. The dynamic states and events taking place in the virtual spaces must be reflected in all processes taking place in these environments. However, coordinating and controlling these states and events can be challenging, especially when several users interact with virtual objects at once without observing any observable lag. In virtual environments, it can be difficult to provide smooth interactions between an infinite number of users at once because any latency might have a negative impact on user experiences. In (Park and Kim 2022), Metaverse hardware components require vision, audio, and tracking systems. Such requirements can be improved using AI technology. Recent AI research work is related to object detection and tracking, action detection, image captioning, NLP, and speech recognition.

3.6 Blockchain technology

The integration of AI with blockchain technology has the potential to enhance the security, privacy, and scalability of Metaverse applications. By leveraging AI algorithms to analyze and learn from blockchain data, developers can create more secure and efficient blockchain-based Metaverse applications. For example, AI can be used to detect and prevent fraudulent activity on blockchain networks, such as identity theft and double-spending attacks (Matthieu et al. 2021). Additionally, AI can be used to analyze and optimize blockchain transactions, improving the efficiency and speed of blockchain networks. By integrating AI with blockchain technology, developers can create more secure and efficient Metaverse applications that offer a range of benefits for users and businesses alike, including increased trust, transparency, and privacy. Moreover, AI and blockchain can provide a platform for developing new and innovative use cases for the Metaverse, such as decentralized virtual marketplaces and secure digital identity systems.

Blockchain is also essential to the growth of the Metaverse since decentralization is a key component of Web 3.0, the next generation of the web. Therefore, even Metaverse initiatives must be developed as decentralized platforms to support a decentralized web. Projects in the Metaverse can benefit from decentralization thanks to blockchain technology (Gadekallu et al. 2022). A network of autonomously operating computers is used to operate the distributed ledger known as the Blockchain. Due to the autonomy of the Blockchain, many parties can conduct transactions among themselves and record those transactions on the blockchain without the need for centralized authority. The Blockchain is transparent and unchangeable, so once anything is recorded it cannot be changed. The blockchain has the benefit of being completely trustworthy. Since blockchain already addresses the problem of double spending, there is no need for confidence between the parties involved in the exchange. In the end, the blockchain resolves any potential fraud or identity theft problems. Blockchain-based prominent Metaverse applications have recently been created (such as the Play to Earn game), and NFTs (Non-Fungible Tokens) play a significant part in these projects. The unchangeable digital proof-of-ownership that is logged onto the blockchain is provided by NFTs.

By incorporating Fungible tokens and NFTs into the Metaverse, many activities relating to the trade of digital goods may be made possible and made easier (Matthieu et al. 2021). Additionally, users might be represented by distinctive Avatars, which would ultimately lead to the development of a Metaverse economy. Users can recreate several behaviors that are similar to our everyday activities in the Metaverse because it is a simulation of the real world. Socializing, having fun, and exchanging are just a few of these activities. The facilitation of a common currency may be necessary for all of these activities, and this can be done by using cryptographic assets. This cryptocurrency can be used by Metaverse users to pay for digital goods and services in both the real world and the virtual world (Gadekallu et al. 2022). The significant role that blockchain has played in the emergence of the Metaverse is examined in (Ning et al. 2021) about the governmental and commercial spheres, while its potential application for VR object connection is stated.

According to (Yang et al. 2022), the traditional Blockchain architecture is composed of several layers, including the data layer, network layer, consensus layer, incentive layer, contract layer, and application layer. The relationship between these layers and the Metaverse is described as follows:

1. The mechanisms for transmitting and verifying data provide network support for various forms of data transfer and verification of the economic system in the Metaverse.
2. The credit problem of Metaverse transactions can be addressed through consensus mechanisms.
3. Blockchain-distributed storage guarantees the safety of digital assets and the identities of Metaverse users.
4. Smart contract technology has created a trustworthy environment for all Metaverse participants. It realizes the value transaction in the Metaverse and ensures that the system rules defined in contract codes are followed transparently. After being launched, the smart contract code cannot be altered. Every clause in these smart contracts must be fulfilled in full.

Combining AI and blockchain is a trend that shows promise for promoting the development of the blockchain/AI-enabled Metaverse ecosystem. The components that blockchains can give for AI through trading in the decentralized market include datasets, algorithms, and computing power. The combination of blockchain, AI, and Metaverse also presents several new research difficulties. Due to the characteristics of digital goods and marketplaces, for instance, transaction volumes in Metaverse systems are far higher than those in the physical world. Avatars can produce material that can be traded with their digital certificates thanks to blockchain-based NFT (Lambert 2021).

The integration of Blockchain technology can strengthen the fundamental technologies that facilitate the Metaverse, enabling users to engage in social and business activities without fear of negative consequences. The Metaverse platform collects data from various IoT devices to ensure seamless integration across a wide range of Metaverse applications, such as healthcare, education, and smart cities. These IoT devices will employ diverse hardware, controllers, and objects to bridge the physical world with the Metaverse. Specific sensors on these IoT devices enable navigation in both the physical and virtual worlds upon connection to the Metaverse.

The ability of IoT devices to perform such tasks is crucial for users to participate in Metaverse activities. Blockchain technology allows IoT devices to communicate across different networks, generating secure records of shared transactions within virtual worlds.

Research in (Deveci et al. 2022) is considering a good example of the power of blockchain with Metaverse to provide a solution for freight fluidity measurement. Three different options are presented in this study for measuring goods fluidity. The technological, governmental, operational, and environmental sustainability merits of these options are assessed. To efficiently and effectively prioritize the options, experts are engaged. The Metaverse application, out of the three options, is the most advantageous to experts. This decision is mostly motivated by how straightforwardly accurate data may be gathered in the Metaverse. Since fluidity measurements are frequently carried out using information about goods transportation, acquiring a lot of trustworthy data using blockchain technology is very promising as a way to improve goods transportation and boost trip time reliability.

With the decentralized administration and control offered by blockchain, users, and applications can access and exchange IoT data, enhancing user trust and preventing conflicts in the Metaverse. All transactions are recorded and verified, ensuring the reliability of data storage in real time. Due to the immutability of blockchain transactions, parties can rely on the information and respond promptly. By enabling users to maintain their IoT data records on public blockchain ledgers, blockchain technology can help resolve issues in the Metaverse (Gadekallu et al. 2022).

The use of AI is essential for the development of the Metaverse and its potential, with AI models creating lifelike Avatars based on user photos or 3D scans. Avatar's attributes affect the user interface, and AI can add dynamic features like facial expressions, emotions, and appearance to the Avatar. Intensive AI training makes the Metaverse accessible to users worldwide, regardless of language ability. Blockchain encryption allows users to control their data and transfer AI ownership to third parties. Zero-knowledge proofs can confirm the accuracy of user information without disclosing it to applications. The audit trail provided by blockchain ledgers ensures transaction authenticity in the Metaverse, while a zero-knowledge evidence system protects privacy and prevents deep fakes. Technical aspects of Blockchain technology in the Metaverse are described in Fig. 6.

From the technical perspective, Blockchain-based methods for the Metaverse can be summarized as follows (Gadekallu et al. 2022):

1. In the Metaverse, it can be difficult to gather reliable data. Blockchain technology can help to get over these restrictions, but because of its distributed nature and complexity, data collection may take longer (Xu et al. 2021). Blockchain transactions may take several days to execute, leading to a limited number of users and higher transaction fees (Alrubei et al. 2020). Moreover, copying data collected on a blockchain along the chain increases the need for storage space, which escalates as data collection in the Metaverse continues to grow (Chen et al. 2022). Thus, the development of mature blockchain technology is necessary to address the challenges in data collection for the Metaverse.
2. The decentralized structure of blockchain technology in the Metaverse can enhance data identification and labeling, as well as promote collaboration among data scientists. Moreover, it offers data availability, reliability, and transparency. Consensus-based distributed ledgers provide enhanced resistance to duplication and tampering, as every block of the blockchain includes a backup of the data (Xie et al. 2019). However, new data must be duplicated throughout the entire chain, leading to delays that require further study. While blockchain technology ensures data immutability, the possibility of a hard fork must be considered. Hard fork refers to the situations when an upgrade in the blockchain process is non-backward compatible with the current blockchain implementation (Yiu 2021).

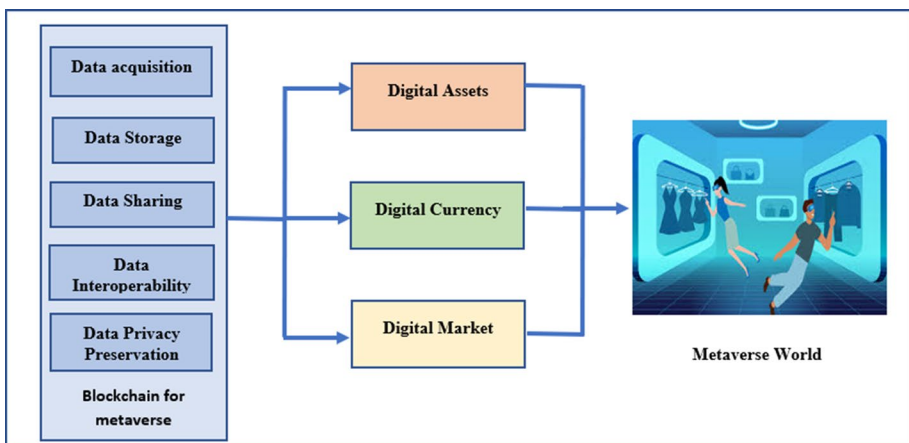


Fig. 6 Blockchain technology for technical purposes within the Metaverse world

3. Data sharing is an essential aspect of the Metaverse, and blockchain technology offers the potential for more adaptable and flexible data management. However, the use of blockchain also poses some challenges. Because each block in the chain duplicates data, transferring information can be slower, and as more users join Metaverse, the computational resources required to process transactions increase significantly. As a result, transaction costs can be higher, which could impede effective data sharing (Luo et al. 2019; Gao et al. 2021). To address these issues, next generation blockchains will need to be developed to improve the scalability and efficiency of data sharing in the Metaverse. Although blockchain technology offers data immutability, the possibility of a hard fork cannot be overlooked.
4. Blockchain technology may enhance interoperability across virtual worlds in the Metaverse, but more study is required. The existence of numerous public blockchains in various VR settings that cannot connect with one another presents the biggest obstacle to cross-blockchain enabled Metaverse interoperability. Because different platforms provide varying degrees of smart contract capabilities, adaptation will be challenging. Additionally, the variety of transaction topologies and consensus processes used by these virtual worlds restricts interoperability (Wibowo and Sandikapura 2019).
5. Maintaining data privacy is of utmost importance in the Metaverse, as a single human error, such as losing a private key, can jeopardize the security of blockchain technology and put data privacy at risk. However, implementing blockchain technology can help protect users' data privacy. In the Metaverse, third-party programs often lack sufficient security measures, which can make personal data susceptible to attacks. Consequently, attackers may focus on these programs (Hassan et al. 2019).

3.7 Digital twins (DT (emerging technology

Operational assets, processes, and systems can be synchronized with the actual world using digital twins (DTs), which are representations of physical objects in virtual space. This encompasses tasks like monitoring, visualizing, analyzing, and making predictions (Tao et al. 2019). Through IoT connections, DTs exist at the nexus of the physical and digital worlds, giving the potential to stop changes made in the real world from reflecting in the digital representation (Chen et al. 2021). DTs are regarded as a crucial component of the Metaverse due to their distinctive qualities and act as a portal for users to access services in the virtual reality. With structure and functionality included, DTs produce exact duplicates of reality that technicians can control using 3D models of complicated systems with varied degrees of sophistication (such as descriptive, instructive, predictive, comprehensive, and so on). DTs can function autonomously and have a wide range of uses, including technical teaching and commercial customization. DTs can be used by developers and service providers to create virtual representations of devices and operations that can be remotely analyzed with the use of AI (Rathore et al. 2021).

The combination of DT and AI is of great importance in the Metaverse world for several reasons:

1. **Realistic Simulation:** DT provides a realistic simulation of real-world entities in the digital world. This simulation can be enhanced with the application of AI algorithms, which can help to replicate complex real-world scenarios with high accuracy.

2. **Personalization:** AI can be used to personalize the user experience within the Metaverse based on data gathered from DT. This can help to create customized services and environments for individual users, enhancing their overall experience.
3. **Predictive Analytics:** AI algorithms can also be used to analyze data gathered from DT to predict future trends and scenarios. This can be particularly useful in identifying potential problems before they occur, allowing for proactive solutions to be implemented.
4. **Autonomous Decision Making:** With the use of AI DT can make autonomous decisions based on the data they collect and analyze. This can be particularly useful in scenarios where real-world entities need to be controlled or managed in the digital world.

Overall, the combination of DT and AI in the Metaverse world allows for a more realistic, personalized, and proactive approach to digital simulation, which can greatly enhance the user experience and improve the efficiency of various systems and processes.

4 Applications of AI in the Metaverse

Metaverse is a persistent, unified network of 3D virtual worlds that will eventually serve as the gateway to most online experiences, it will also underpin much of the physical world. These concepts, which were formerly restricted to science fiction and video games, are now being developed to revolutionize every sector of the economy, from finance and healthcare to education and consumer goods. This section provides a deep discussion about how we can create a Metaverse world using different AI technologies.

4.1 Avatar creation and customization

Avatar creation and customization involve the process of constructing a 3D virtual world to provide users with an immersive experience. The process of creating immersion involves creating virtual scenes in the Metaverse and displaying them to users, using inspiration from computer vision, graphics, and visualization techniques (Zawish et al. 2022). The Metaverse's virtual world combines physical and digital elements and is made up of a variety of settings, NPCs, and player characters (avatars). Scenes show various virtual locations, including a campus (Duan et al. 2021) or a museum (Beer 2015). NPCs are objects that cannot be controlled by the player, but they play a crucial role in making the game world feel more lifelike (Warpefelt and Verhagen 2015). Contrarily, avatars are virtual representations of participants in the Metaverse that allow for communication with other participants or AI agents (Davis et al. 2009). Users can engage with others in a real-world setting without any physical restrictions by becoming a Metaverse Avatar. Computer vision techniques enable XR devices to recognize and comprehend visual data about user activities and their physical surroundings, enabling the creation of these worlds and the interaction between objects, resulting in more accurate and precise virtual and augmented environments. Figure 7 illustrates the process of creating immersion in the Metaverse.

Avatar creation in the Metaverse is far more user-defined than the other two components, with users contributing to numerous features in 3D modelling and animation development. Users often have a variety of options to change and personalize their avatar's appearance. Most video game designers either employ a small number of models or allow players to construct whole Avatars with just a few optional sub-models, like the nose, eyes, mouth, and so forth, resulting in player Avatars that are all identical

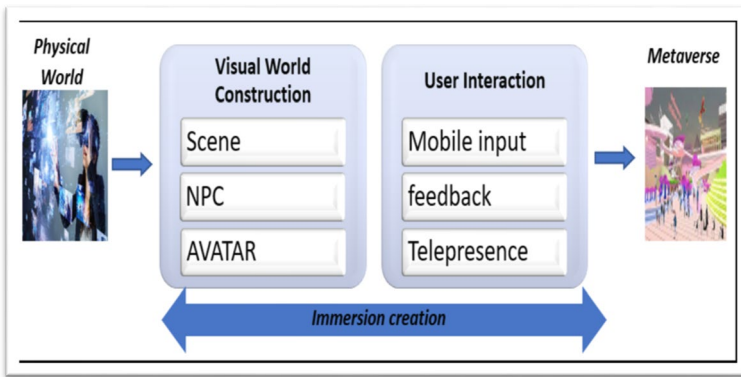


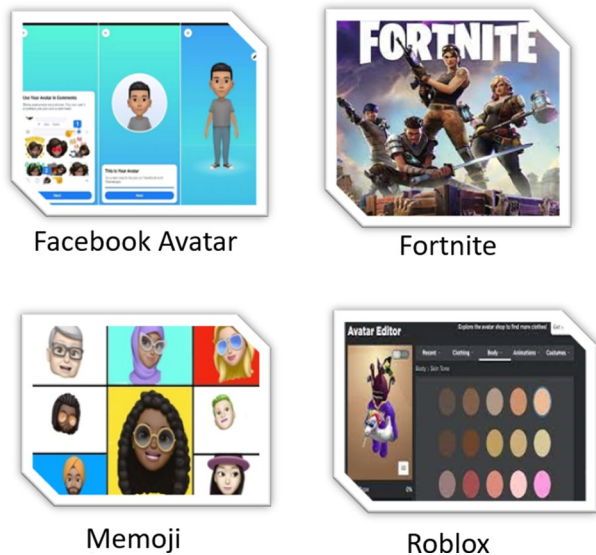
Fig. 7 The immersion creation for Metaverse

in appearance (Lee et al. 2021). Certain programs allow users to scan their physical appearance and choose their virtual attire to imitate their real-life appearances. Despite a major improvement in realism, design-based avatars still retain a cartoonish appearance. In the Metaverse, during various social interactions, specifics like the Avatar's facial features (Wei et al. 2004) and micro-expressions (Murphy 2017), as well as the entire body (Kocur et al. 2020), can impact user perceptions. Therefore, to enhance realism, various reconstruction technologies have been developed to create highly realistic 3D faces and bodies.

The process of reconstructing faces plays a crucial role in the creation of Avatars. Conventionally, this process is carried out using 3D Morphable Models (Booth et al. 2016) based on the Principal Component Analysis algorithm, which has limitations in capturing facial details due to data quality. Deep learning approaches, such as Generative Adversarial Networks (GAN) and SDF (Yanghua et al. 2017; Hongyu and Tianqi 2019; Koichi et al. 2018), have, however, recently appeared and have attained a higher level of realism. Some researchers have utilized GAN to produce 2D avatars for video games, while others have produced 3D avatars by instantly processing 3D meshes and textures. An application for creating autonomous 3D avatars has been presented by authors in (Chai et al. 2016; Takayuki and Takashi 2019; Tianyang et al. 2019) using face scanning rather than 2D images (Igor et al. 2017). In addition, StyleGAN (Luo et al. 2021) has been used to create highly photorealistic 3D face templates, as demonstrated in (Luo et al. 2021). Recently, Microsoft proposed "Rodin", a 3D generative model that generates 3D digital Avatars represented as neural radiance fields using diffusion models (Wang et al. 2022).

The study of 3D bodies has been extensively researched, including the micro-neural rendering of 3D bodies based on different data types (Wu et al. 2020) and the reconstruction of high-realistic human dynamic geometry and textures from simple input (Liu et al. 2021). However, due to their small size, hands are frequently obscured when scanning or photographing whole bodies, which presents a barrier for researchers. To address this, a number of models have been put forth to depict human hands, the most well-known of which being MANO (Romero et al. 2017), which is well-suited for deep learning networks. Typically, user interaction with controllers or real-time tracking is used to create avatar animation (Genay et al. 2021). Real-time tracking technologies

Fig. 8 Several real-life examples of Avatars



have become essential for generating animations since user interaction alone may not create strong ownership illusions through inputs that do not represent the user's face, hands, or body. By mapping user movements, these technologies allow for more accurate control of Avatars.

Several real-world examples of avatars are shown in Fig. 8, including Facebook Avatar, which lets users customize their social network avatars, Fortnite, a multiplayer game that lets players build, create, and customize their virtual worlds, VR Chat, a VR experience, and Memoji, which exemplifies a type of augmented reality and enables users to interact with cartoonized faces during FaceTime on Apple iOS devices. Avatars can also be human-like ones that resemble real humans. We have created an Avatar dataset (Ahmed et al. 2023) that comprises 130 human-like avatars of real persons that differ in shape, color, gender, and age. Regarding gender, there are 70 avatars for males and 60 avatars for females. Regarding age, there are 13 avatars for old people ages 50 to 70 years old, 94 avatars for adults ages 18 to 49 years old, and 23 avatars for children. We used Avatar SDK (Avatar Sdk, <https://avatarsdk.com/>) which converts selfies into realistic avatars. The dataset consists of an image of the avatar and a video that captures some facial movements ranging from 15 to 25 seconds for every person. Examples of the generated dataset are depicted in Fig. 9.

We use this dataset for authenticating avatars in the Metaverse. Avatar authentication is necessary for controlling the roles of users in the Metaverse.

4.2 Metaverse world construction

Metaverse visual world construction embraces different technologies including Augmented Reality, Virtual Reality, Extended Reality, and mixed reality and their corresponding tasks. The majority of VR, AR, and XR systems use an optical see-through or video see-through display to record visual data. A head-mounted device or a smartphone is used to process this data and deliver the results. To process, evaluate, and grasp visuals like digital photos or movies make decisions, and take necessary action, the use of computer vision is



Fig. 9 Examples of the generated Avatars using OpenCV

inevitable. It is also related to computer-vision-based methodologies. Visual world Creation tasks which are critical to the success of every Metaverse application can be divided into three main parts: scene generation and recognition, Non-player character (NPC), and Player character (avatar) construction.

Users will encounter numerous vibrant and realistic scenes in the Metaverse. The construction of the scene places a greater emphasis on the realism of architecture and layout. Mesh models are more appropriate for reconstructing some buildings characterized by complex surface shapes like murals and statues. Leotta et al. (Leotta et al. 2019) for instance, used mesh models and segmentation techniques to reconstruct urban buildings. Navarro et al. used the mesh method to reconstruct indoor rooms into virtual worlds (Navarro et al. 2017). Some researchers looked into how to enable real-time reconstruction access for remote multiple clients in a virtual environment. Making physical-based constructions is one method for creating 3D models. This is accomplished by using 3D measurement techniques, photogrammetry and notably laser scanning to build digital twin models (Deng et al. 2021). In VR and AR, there is currently an increased demand for producing high-quality and detailed photorealistic 3D content based on real objects and environments (Zhao et al. 2022). Building multi-scene graphs is proposed in (Zhu 2022), this framework merges many scene graphs into a single graph to allow for more thorough analysis and

inference in addition to describing rich interactions in a single scene. SLAM Cast developed a practical client–server system for real-time capture and multi-user exploration of static 3D scenes (Stotko et al. 2019).

Rendering is a crucial step for providing virtual objects with real features and presenting them in front of our eyes. Rendering includes lightening, coloring, and texturing. Because many applications demand color information, laser scanning does not provide it. To colorize the point cloud data, a hybrid 3D reconstruction based on scan data and photos is applied. Ma et al., for instance, presented a differential framework for capturing freestyle material (Ma et al. 2021). This technique promotes the production of more realistic “virtual things” in the Metaverse by significantly increasing sampling efficiency and more specifically restoring material properties. Many modeling programs are available, and they can be categorized into three types based on how they implement modeling and rendering. First, engineers and designers choose to use parametric 3D modeling, also known as CAD (computer-aided design) software, to construct models by setting the parameters like the real thing: materials, weight, etc. Second, polygon models are more concept-oriented than measurement-oriented, in contrast to CAD modeling. The popular software includes Blender, Maya, and 3Ds Max. Third, compared to polygon modeling, using digital sculpture modeling tools like ZBrush6 demands more artistic talent. To create the virtual world, modeling software can be integrated with engines (such as Unity3D or Engine). Computer graphics software uses design-based animation to provide the appearance that objects are moving in 3D space. The scene generation process has accelerated greatly to almost real-time. However, the scenes produced by these computer-vision-based methodologies are constrained by the predefined elements used in AI algorithms, prohibiting the immersive experience of human connection with a digital human in the Metaverse because these aspects aren't the same as actual ones. Yet we want both parties to share the same scene and, consequently, the same environment. This makes the Metaverse's sensing, sampling, and scene generation more difficult. One option is to record actual scenes and then utilize those pictures or films to create new ones, updating the scenes in the Metaverse quickly, possibly in real-time (Zawish et al. 2022).

Scene recognition is another solution by using an image recognition task, which entails predicting a location, based on some pre-defined categories. Scene recognition is very crucial to achieving successful interactions and harmony between applications in the Metaverse world. There is previous and ongoing research on Scene recognition (Deng et al. 2021). Deep learning (Zawish et al. 2022; Stotko et al. 2019) has been widely used in the field of scene recognition. Especially, CNN has been extensively studied for different problems, all aimed at achieving an effective scene recognition process (Lin et al. 2021; Deng et al. 2021). In (Deng et al. 2021), two different CNN networks are used to extract global and local visual features resulting in more robust scene recognition. CNNs have also been used for getting low-level features, to decrease the prediction time of complex scene recognition (Seong et al. 2020). In (Lee et al. 2021), ImageNet, a popular object detection dataset, was fine-tuned using a deep CNN, to train a scene recognition dataset. Some researchers have gone further to employ fusion techniques, by fusing objects and scene information using CNN (Seong et al. 2020). In virtual spaces such as Metaverse, the problem of scene recognition has not yet been explored. Two Convolutional Neural Networks (CNN) models: SimpleNet and AlexNet, are applied in (Deng et al. 2021) to automatically recognize Virtual scenes. A real-world scenes dataset called Scene15 is used for training then the model is tested on virtual-world scenes. From the results analysis, the AlexNet model performed much better than the SimpleNet model in terms of training time, test time, training accuracy,

and test accuracy. This is mainly because it was already pre-trained for image classification tasks. Another way to help scene recognition is holistic scene understanding. We gain an understanding of the physical world by responding to four basic questions: what is my role? What's in my surrounding environment? How close am I to the mentioned object? What could the thing be doing? These questions are addressed by comprehensive scene understanding in computer vision. We must communicate with users and other objects in both the real and virtual worlds because of the Metaverse. Accordingly, holistic scene understanding plays a crucial role in ensuring the operation of the Metaverse. (Lee et al. 2021) Computer vision based on artificial intelligence approaches can help in providing solutions to the scene understanding problem. Semantic Segmentation and Object Detection are among the proposed solutions. Semantic segmentation is a computer vision task that divides an image into categories depending on the information contained in each pixel. (Lin et al. 2021; Lin et al. 2021), It is considered one of the key techniques to completely understand the environment (Tanzi et al. 2021). In computer vision, a semantic segmentation algorithm should efficiently and swiftly segment each pixel based on the class information. Recent deep learning-based methods (Lin et al. 2021; Lin et al. 2021), and (Liang-Chieh et al. 2018) have demonstrated a notable improvement in performance in urban driving datasets intended for autonomous vehicles. Accurate semantic segmentation in real-time, however, continues to be difficult. For instance, semantic segmentation algorithms need to execute at a pace of about 60 frames per second (fps) for AR applications (Ko and Lee 2020). Semantic segmentation is therefore an essential but challenging task for realizing the Metaverse.

Another key task for understanding a scene is object detection, which aims to localize the items in it and determine the class information for each one (Luyang et al. 2019). XR makes extensive use of object detection, which is crucial for realizing the Metaverse. For instance, a typical object identification task in VR is face detection, whereas a typical object detection task in AR is word recognition. A more advanced use of AR object recognition tries to integrate a 3D representation into the real world. To accomplish this, object detection algorithms must accurately determine the type and location of objects. Users can move and modify a physical object by positioning a 3D virtual object nearby and attaching it to it. The Metaverse's 3D world can be made richer and more immersive with the aid of AR object identification. The early attempts at semantic segmentation primarily incorporate feature tracking techniques, such as SIFT, that seek to classify the hand-crafted features, like the support vector machine (SVM), before segmenting the pixels. Both VR and AR have used these algorithms (Jenny et al. 2016; Peer et al. 2019). These traditional techniques, however, have poor segmentation performance. Recent studies have investigated CNN's potential for semantic segmentation. These methods have been successfully applied to AR (Tanzi et al. 2021; Liang-Chieh et al. 2018; Huanle et al. 2020). The ability of semantic segmentation to address the occlusion issues in MR has been demonstrated in certain publications (Kido et al. 2021; Menandro et al. 2018). Image segmentation causes a significant computational and memory load because it employs each pixel, though. To comprehend the pixel-wise information in a 3D immersive environment in the Metaverse, we need real-time and reliable semantic segmentation approaches. The diversity and complexity of virtual and actual objects, contents, and human avatars necessitate the development of more flexible semantic segmentation approaches. The semantic segmentation algorithms must be able to tell the real pixels from the virtual ones, especially in the inter-laced Metaverse world. In this situation, the class information may be more complex, and the semantic segmentation models may be required to work with classes that are not yet known.

There are two types of object detection in the Metaverse: detection of unique instances (such as a face, marker, or text) and identification of general categories (such as automobiles and people). In XR, (Hbali et al. 2016; Nahuel et al. 2020), text detection techniques have been extensively researched. These techniques are already developed and can be used immediately to create the Metaverse. In recent years, there has also been a lot of research on face detection, and the techniques have proven to be reliable in a variety of recognition settings for XR applications, such as (Xueshi et al. 2020; Tanja et al. 2020; Amin et al. 2020; Bernardo et al. 2019; Jan and Jonatan 2018). In the Metaverse, users are represented by avatars, and these avatars can communicate with one another. Both authentic faces from the real world and artificial faces from the virtual world must be detected by face detection algorithms. Additionally, it may be more difficult to identify faces in the Metaverse due to occlusion issues, sudden changes in face posture, and illumination variations. The research community has focused a lot of attention recently on the detection of generic categories using deep learning. One of the SoTA techniques in the early stages of deep learning development was the two-stage detector, FasterRCNN (Ren et al. 2015). One of the SoTA techniques in the early stages of deep learning development was the two-stage detector, FasterRCNN (Ren et al. 2015). After that, the Yolo series and SSD detectors (Joseph and Ali 2018; Alexey et al. 2020; Wei et al. 2016) revealed excellent detection performance on different scenes having multiple classes. Successful applications of these detectors have been made to AR (Sagar 2018; Martin et al. 2018; Haythem et al. 2019). It has already been demonstrated that SoTA object detection techniques are effective for XR. To go to the Metaverse, there are still several challenges to overcome. The detection of smaller or tiny objects presents the first challenge especially when multiple things co-exist in the same space in the 3D immersive world. Some items and objects will get smaller as the camera's Field of View (FoV) changes, making them difficult to see. Some items and objects will get smaller as the camera's Field of View (FoV) changes, making them difficult to see. To detect these items independent of the capturing tools, the Metaverse's object detector needs to be strengthened. Issues with data and class distribution make up the second. In the Metaverse, it is often simple to gather large datasets with more than 100 classes, but it is more difficult to gather datasets with a diverse scene and class distribution. The final one is the Metaverse's object detection computational burden. The Metaverse's 3D immersive universe has a variety of information that must be distributed even in remote areas. The amount of computation required increases as the class size grows. To this goal, the research community anticipates the development of more effective and portable detection techniques (Lee et al. 2021).

4.3 User interactivity

The Metaverse offers users, the digital world, and the physical world sophisticated, multimodal interaction services (Yogesh et al. 2022). The main obstacles to ensuring immersive experiences are real-time data flow and multi-player interaction across various Metaverse zones. First, although more intensive rendering computations are needed, intelligent devices facilitate the powerful perception and interaction capabilities of the Metaverse. For instance, real-time interaction, synchronization, and large-scale reconstruction and rendering provide challenges for multi-player interactions. Second, users must synchronize and switch services seamlessly between the physical and digital worlds in various Metaverse zones to employ diverse Metaverse services. Lastly, in a large-scale, complicated situation, flexible setup and optimization of multidimensional

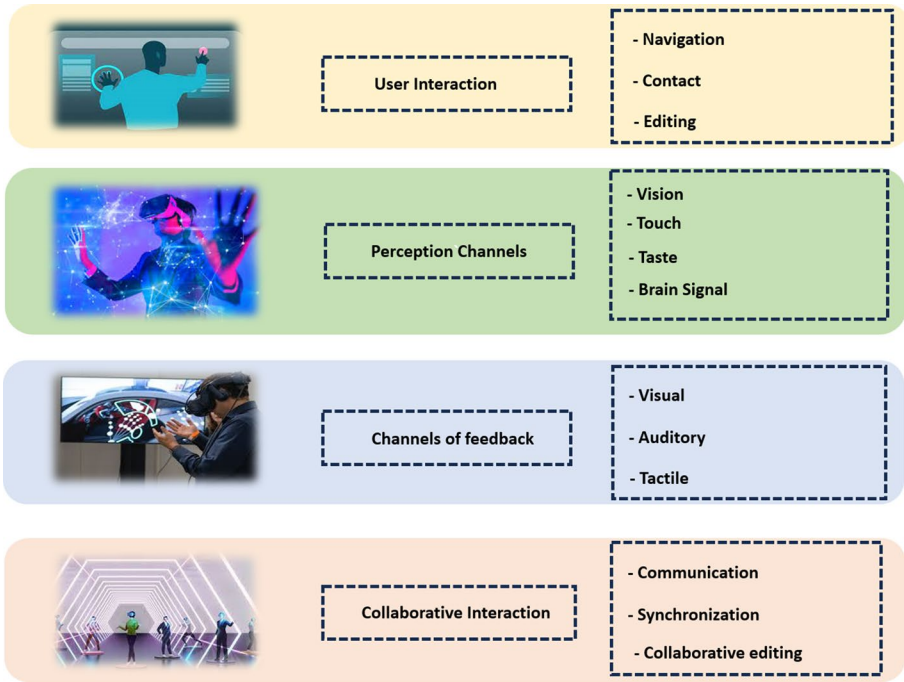


Fig. 10. Main components of interaction service in the Metaverse

network resources are also important for multi-player engagement (Jankowski and Hachet 2013; Yakun et al. 2022). Figure 10 shows the main components of interaction service in the Metaverse.

For a particular interaction task, the user can obtain seamless interactions in various Metaverse zones using physiological signals as inputs and terminals for feedback output. Navigation, contact, and editing are the three operations that make up the general interaction tasks (Zhao et al. 2022). Geographic and non-geographic navigation are both included in interaction navigation, which is the user's current view operation. Geographic navigation is the movement of a user across the real world or the virtual world utilizing an interaction terminal. By asking a question, performing a task, or taking other specific steps, non-geographic navigation completes the navigation process (Jankowski and Hachet 2013). Contact interaction is the use of multiple sensors, including direct and indirect ways, to touch and sense the physical world. For instance, in the direct touch technique, people can experience the physical world quite similarly by using haptic and force feedback sensors to observe it. Indirect contact means that users can manipulate virtual physics in the digital environment. Any dynamic or entity that can be altered in the digital world is subject to editing interaction. There are numerous perceptual channels available for users to interact with others in the Metaverse. Brain signals, taste, touch, and vision. A variety of signals, including electromyography (EMG), electroencephalography (EEG), biopotentials, and many others, are gradually coming to be understood as potential interactive control inputs for the Metaverse that could free up user participation. Additionally, the visual display or sensors of the

intelligent interaction devices can provide feedback to people and their surroundings. Visual, auditory, and tactile modalities are only a few of the feedback channels for engagement. Users can engage with virtual representations in the digital world more realistically and receive immediate visual and acoustic input, for instance. Several force feedback sensors, such as smart XR gloves that can simulate information about the user's shape and force, such as stiffness, friction, and sliding, can provide haptic feedback (Yakun et al. 2022). Collaborative interaction, which consists of communication, synchronization, and collaborative editing, is another more significant form in the Metaverse. For instance, users must communicate and synchronize the information of various users in real time when they are in the same Metaverse region. Also, collaborative editing is needed when users interact across various Metaverse zones. The employment of user interaction approaches in both physical and virtual contexts has recently been proposed. Young et al. (Mary et al. 2015) suggested an interaction paradigm that allows users to synchronize high-fiving gestures in both real-world and virtual contexts. An interactive system was suggested by Vernaza et al. to link smart wearables (Vernaza et al. 2012). The customization of virtual characters in virtual worlds was then made possible by Wei et al.'s user interfaces (Yungang et al. 2015). Third, the research community is beginning to pay more attention to user activity analysis in the Metaverse. The text content produced in many virtual worlds (Gema and David 2013) and avatar actions in virtual settings (Gema 2012) may be understood using well-known clustering techniques (Gema and David 2013). Since the Metaverse may connect the users with other non-human animated objects, a study by Barin *et al.* (Amirreza et al. 2017) investigates the crash incidence of high-performance drone racing through the first-person view of VR headsets. The study's conclusion statement encourages that user-drone interaction in virtual settings will no longer be constrained by physical factors like acceleration and air resistance. Instead, the layout of user interfaces could slow down users' responses and cause serious crash issues.

4.4 Text/audio understanding

The Metaverse is marketed as a place without boundaries and a virtual setting where individuals may be themselves without having to consider their racial or linguistic background. It's simpler to detect bad behavior when technology can quickly grasp and interpret natural language. Users may be removed from platforms when this information is identified, protecting other users. Furthermore, the system may even be able to recognize people who may be susceptible to such attacks and initiate preventative support measures for them. Natural language processing (NLP) techniques will enable this by instantly decomposing natural languages and converting them into a machine-readable format (Güven and Balli 2022). This machine-readable text will undergo processing before being translated back into another language. Edge computing (Khan et al. 2019) can accomplish all of this in a matter of seconds, giving the conversion a realistic appearance even in the virtual world. Facebook recently unveiled Horizon Workrooms (Meta 2021), a well-designed meeting platform which enables users to work, collaborate, and connect with others in the virtual world using VR devices in addition to conducting training and coaching sessions. Users are represented by avatars on this platform. AI avatars could be used by businesses as assistants to guide consumers through virtual spaces for a better user experience. For enterprises joining the Metaverse, this level of human-like communication will have far-reaching effects. They'll be in a better position to comprehend and interact with customers. They

will be able to offer great, individualized client service around the clock. Additionally, by outsourcing tasks to skilled AI service providers, they will significantly reduce labor costs (<https://accelerationeconomy.com/Metaverse/how-nlp-can-enhance-the-Metaverse-experience/>). Chatbots with AI are becoming more and more common in businesses. In the Metaverse, AI bots can be employed for a range of purposes, including sales, marketing, and customer support. It can assist users in the Metaverse by giving directions, offering details about goods and services, responding to inquiries from clients, taking orders, and even carrying out transactions on their behalf. Meta announced a next-generation AI model called Project CAIRaoke (<https://ai.facebook.com/blog/project-cairaoke/>) that will power chatbots and assistants in its Metaverse.

To completely support text- and speech-based interactive experiences between human users and virtual assistants in the Metaverse, NLP approaches should be merged. The major method of communication between entities (avatars, virtual people, or even non-human things) in the Metaverse is voice, which is seen as one of the key interfaces for humans to enter the Metaverse. The two primary tasks involved in voice processing are automatic speech recognition and text-to-speech (or speech synthesis), which converts voice signals to text or vice versa. In the Metaverse, along with language understanding tools, automatic speech recognition and text-to-speech enable entities to understand the message and purpose of others and to communicate as if they were in the real world. Using NLP, digital objects within the Metaverse can also “speak back”. VR users frequently receive responses from non-player characters (NPCs) or artificial humans in the form of speech bubbles. These interactions would be elevated to a whole new level by NLP, which would enable the production of audio answers with voice modulation and subtle grammatical variations. To reach a larger audience, it might even automatically convert the response into several languages.

Additionally, audio signals might be converted into binaural signals so that people might experience auditory immersion and a feeling of the position of the sound source in an enclosed area. Many obstacles still need to be overcome to achieve full auditory immersion in the Metaverse. These obstacles include distinguishing between Metaverse and real-world sound, producing customized speech synthesis, creating complex 3D soundscapes, etc., all of which heavily rely on cutting-edge signal processing and machine-learning techniques. The Metaverse is filled with sound, and we think AI is playing a major role in audio management there. In the following, we illustrate in more detail some key concepts in NLP that can be used to create and enhance a Metaverse world.

5 Problems and challenges

5.1 Avatar creation and customization

There are some of the challenges and problems associated with using AI for Avatar creation and customization in the Metaverse world:

1. **Quality of generated Avatars:** One of the primary challenges with using AI for Avatar creation is generating high-quality Avatars. This is because current AI models may not have the ability to create highly detailed, realistic, and unique Avatars, which may result in a lack of diversity in the Metaverse.

2. **Personalization:** Personalization is a key aspect of Avatar creation, and it is challenging to train AI models to generate highly personalized Avatars that reflect the user's personality, interests, and preferences.
3. **Data privacy and security:** The collection and use of user data to create Avatars and customize them can raise privacy and security concerns, especially if the data is not adequately protected.
4. **Accessibility and inclusivity:** In the Metaverse world, Avatars are essential tools for users to interact with each other. However, not all users may have equal access to AI-generated Avatars, and this can result in the exclusion of certain groups of people.
5. **Ethical considerations:** AI-generated Avatars can create a risk of perpetuating stereotypes, biases, and discriminatory practices if the training data used to create the models is not diverse or representative of all groups of people.
6. **Continual learning:** AI models for Avatar creation must continually learn and evolve to meet the changing preferences and needs of users. This requires a significant number of resources and ongoing maintenance to keep up with the demands of the Metaverse world.

Overall, while the use of AI for Avatar creation and customization offers many benefits, such as increased efficiency and personalized experiences, there are significant challenges that must be addressed to ensure that this technology is used responsibly and ethically.

5.2 Computer vision with AI in the Metaverse

Some of the challenges and problems associated with using computer vision with AI in the Metaverse world are presented in the next:

1. **Realism and accuracy:** One of the primary challenges of using computer vision with AI in the Metaverse is achieving a high level of realism and accuracy. This is because the Metaverse is a virtual world, and the images and video captured by computer vision algorithms must accurately represent the virtual environment.
2. **Computational power and scalability:** Computer vision algorithms need significant computational power to process large amounts of data quickly. As the Metaverse continues to grow, there may be challenges in scaling up computer vision algorithms to handle the increasing amounts of data.
3. **Privacy and security:** The use of computer vision with AI in the Metaverse raises concerns regarding privacy and security. This is because the algorithms may capture and process sensitive information about users without their knowledge or consent.
4. **Bias and fairness:** Computer vision algorithms can be biased and unfair if the training data used to develop the algorithms is not diverse and representative of all groups of people. This can lead to discriminatory practices and exclusion of certain groups of people in the Metaverse.
5. **Transparency and interpretability:** The complexity of computer vision algorithms makes it challenging to figure out how they work and how they make decisions. This lack of transparency and interpretability can make it difficult to identify and correct errors or biases in the algorithms.
6. **Environmental impact:** The computational power required for computer vision algorithms can have a significant environmental impact, especially if the algorithms are run on energy-intensive hardware.

Overall, while the use of computer vision with AI in the Metaverse offers many benefits, such as improved user experiences and enhanced interactions, there are significant challenges that must be addressed to ensure that this technology is used responsibly and ethically.

5.3 Natural language processing with AI in the Metaverse

Some of the problems and challenges of using NLP with AI in the Metaverse world are highlighted in the following:

1. **Understanding context:** Understanding the context of the text is one of the biggest challenges in NLP. In the Metaverse, users may use slang, jargon, or other forms of language that are not commonly used in the real world. This can make it challenging for NLP algorithms to understand the meaning of the text accurately.
2. **Multilingual support:** The Metaverse is a global platform, and users from different parts of the world may use different languages. NLP algorithms must be able to support multiple languages to enable effective communication between users.
3. **Real-time processing:** In the Metaverse, users may communicate with each other in real-time. This requires NLP algorithms to process text quickly and accurately to facilitate seamless communication between users.
4. **Privacy and security:** NLP algorithms may collect and process sensitive information about users, such as their personal preferences, interests, and behavior. This can raise privacy and security concerns, especially if the data is not adequately protected.
5. **Bias and fairness:** NLP algorithms can be biased and unfair if the training data used to develop the algorithms is not diverse and representative of all groups of people. This can lead to discriminatory practices and the exclusion of certain groups of people in the Metaverse.
6. **User experience:** The success of NLP algorithms in the Metaverse depends on how well they can provide a seamless user experience. This includes ensuring that the algorithms can understand user requests accurately, provide relevant information, and respond quickly.

Overall, NLP with AI offers many benefits in the Metaverse world, such as facilitating communication between users and providing personalized experiences. However, there are significant challenges that must be addressed to ensure that this technology is used responsibly and ethically.

5.4 Legal and ethical considerations in AI-enabled Metaverse applications

This subsection addresses the legal and ethical considerations of using artificial intelligence (AI) in the Metaverse. The integration of artificial intelligence technologies within the Metaverse raises important legal and ethical considerations that warrant attention. In this subsection, we explore key aspects of the legal and ethical landscape surrounding the use of AI in the Metaverse.

- **Intellectual Property Rights:** The utilization of AI algorithms and models within the Metaverse necessitates an examination of intellectual property rights. Issues may arise in terms of copyright, patentability, and ownership of AI-generated content. Questions

regarding the ownership of AI-generated artworks, virtual goods, and other digital assets may require clarification within existing intellectual property frameworks.

- **Privacy and Data Protection:** AI technologies in the Metaverse often rely on extensive data collection and analysis. The collection, storage, and processing of personal and behavioral data in virtual environments raises concerns related to privacy and data protection. The legal landscape surrounding data protection and user consent must be considered to ensure compliance with relevant regulations such as the General Data Protection Regulation (GDPR) and similar legislation.
- **Liability and Accountability:** The dynamic nature of the Metaverse and the autonomous decision-making capabilities of AI systems give rise to questions of liability and accountability. In the event of AI-related errors, accidents, or malicious activities within virtual environments, determining responsibility and assigning liability can be challenging. Legal frameworks should address these issues to ensure appropriate accountability and the protection of user rights.
- **Regulatory Frameworks and Governance:** Given the potential societal impact of AI-enabled Metaverse applications, the development of comprehensive regulatory frameworks becomes crucial. Governments and regulatory bodies should establish guidelines that govern the use of AI in the Metaverse, including standards for safety, security, and ethical considerations. Striking a balance between innovation and protecting public interests will be pivotal in shaping the future regulatory landscape.
- **Ethical Concerns and Human Rights:** Ethical considerations surrounding AI and the Metaverse extend beyond legal compliance. Questions of fairness, transparency, and inclusivity arise concerning AI-generated content, virtual experiences, and social interactions. The potential for discrimination, bias, and the erosion of human agency within virtual environments must be critically evaluated, and ethical guidelines should be developed to mitigate these risks.

By addressing these legal and ethical considerations in the context of AI-enabled Metaverse applications, we can foster responsible development and deployment of AI technologies within virtual environments. A comprehensive understanding of the legal and ethical landscape will contribute to the creation of a more inclusive, secure, and ethically sound Metaverse ecosystem.

6 Recommendations and future directions

A virtual transformation will be driven and integrated by AI and Metaverse. Furthermore, technical professionals are interested in the Metaverse and its future. A VR world called Metaverse allows for user interactions using a variety of technologies, including AI, AR, VR, etc. Users can interact with 3D digital items and virtual Avatars utilizing a variety of tools and technologies. Additionally, AI and Metaverse work together to create innovations and discoveries that usher in a new era of reality.

AI is widely used across many applications, including ML algorithms and deep learning structures. These techniques have demonstrated significant value in a variety of areas. For example, ML methods such as unsupervised, semi-supervised, supervised, and reinforcement learning have been applied to classification and regression models for language-processing tasks, like voice recognition, which enables system agents to understand user commands. Additionally, complex patterns of human actions can be analyzed and learned

through physical activity recognition, allowing systems to comprehend user activities and interactions in the virtual world by gathering sensor-based signals from multiple devices, including mobiles, smartwatches, and other wearable devices. Deep Learning architectures such as recurrent neural networks (RNN), long short-term memories (LSTM), and convolutional neural networks (CNN) have emerged as powerful AI techniques for identifying complex patterns from large datasets and have achieved great success in the computer vision industry. Deep Learning is now being used in various industries, including gaming, human-computer interaction, wireless communications, and finance.

The use of AI is crucial to ensure the reliability and performance of the infrastructure in the Metaverse, as demonstrated by the seven-layer Metaverse platform. The value chain of the Metaverse market is described by these seven layers. Opportunities, technological advancements, and solutions fall under this category. Without the base that infrastructure provides, none of the next advancements would be possible. The Metaverse is therefore driven by technical procedures. The other layers include experience, discovery, the creative economy, spatial computing, decentralization, and human interface. The experience layer: The excitement and resources the Metaverse has attracted are directly related to the realistic experiences it is positioned to offer. The Metaverse is made up completely of experiences. Gaming, social interactions, eCommerce, entertainment, and e-sports may all be revolutionized by the immersive and real-time features of a true Metaverse.

- The discovery layer: This layer describes how app stores, search engines, review websites, and display advertising can help users find new experiences or platforms. Most discovery tools fall into one of two categories: inbound or outbound. Finding new platforms, protocols, and communities is an essential stage in the development of Metaverse technology.
- The creator economy: To create digital resources, immersive experiences, and other assets, developers and content producers use a wide variety of design tools and software. As time goes on, an increasing number of systems offer drag-and-drop functionality to make creating easier. As Web3 grows more established in culture and Web2 is gradually phased out, becoming a maker, developer, or designer has never been easier. This is evident on numerous Metaverse platforms, such as The Sandbox, which makes the creation of digital assets incredibly straightforward and code-free.
- The spatial computing layer: A technology solution known as spatial computing combines virtual and augmented reality (VR/AR) to deliver a high level of authenticity. Spatial computing gives users the ability to navigate 3D environments, digitize objects using the cloud, and interact with their surroundings via spatial mapping, which displays information related to actual areas in user contexts.
- The decentralization layer: The Metaverse would be controlled by a decentralized autonomous organization (DAO) with open ownership and would be decentralized, open, distributed, and decentralized. Normal users would be unable to tell who had access to it and under what circumstances due to central ownership. Users would be dissatisfied if there were security lapses as a result of this. A revolutionary solution to the privacy and data security issues that could arise in a centralized Metaverse is blockchain technology. Decentralized apps (dApps), also known as blockchain-based applications, are being created and utilized in a wide range of industries to make use of the inherent security and decentralization of the blockchain.
- The human interface layer: This layer describes the HCI (very advanced human-computer interaction) technologies that enable users to explore the Metaverse. Hence, it includes haptic technology, smart eyewear, and VR headsets that let users move around

virtual worlds. People will be able to learn more about their environment by using products like Google Glass or Project Aria from the Meta Platform.

- The infrastructure layer: The technologies in the seventh help in achieving earlier concepts in reality. 5G-capable infrastructure is important to boost network capacity and reduce congestion and latency. To function properly, the devices found in the human interference layer also need components such as tiny, long-lasting batteries, semiconductors, and microelectromechanical systems (MEMS). Wi-Fi, blockchain, artificial intelligence (AI), cloud computing, and graphics processing units are among the technologies that support the Metaverse.

In the 5G and upcoming 6G systems, advanced supervised learning and reinforcement learning ML algorithms have been implemented for a range of challenging tasks, including effective spectrum monitoring, automatic resource allocation, channel estimation, traffic off-loading, attack prevention, and network fault detection. Additionally, wearable sensor-based technology and other devices have been used with ML and Deep Learning models for human-machine interaction, allowing for the recognition of simple gestures and complex activities. As a result, users' physical movements can be translated into actions in the virtual world, giving them complete control over their avatars' interactions (Yakun et al. 2022). Moreover, AI enhances the accuracy and processing speed of voice recognition and sentiment analysis, which includes facial expressions, emotions, body movement, and physical interactions, enabling avatars to interact with various real-world modalities.

Although head-mounted displays are used in XR/VR to create immersive experiences, AI plays a crucial role in the background to enhance the virtual world's appearance and provide a seamless user experience. AI can simplify the content creation process, with tools such as NVIDIA's GANverse3D that can transform photographs into virtual reproductions. Various deep learning-based techniques can achieve high accuracy and real-time processing for displaying 3D objects, using software such as Facebook's PyTorch3D library and hardware. Meta's AI research supercluster, which is one of the fastest AI supercomputers globally, is designed to speed up AI research and support the development of the Metaverse (Yogesh et al. 2022). The supercluster can aid AI scientists and researchers in creating stronger deep learning models from large data sets, such as text, speech, image, and video, for various applications and services. All the supercluster's achievements and outcomes will be integrated into the Metaverse platform to create AI-driven goods and promote user interaction with virtual assistants and other users. AI also broadens the Metaverse's potential, allowing individuals and organizations to create, acquire, and market diverse goods, services, and solutions. This encourages collaboration and communication between people and companies for better opportunities. Advanced technologies such as 5G, XR/VR, and blockchain are combined to create a fully 3D immersive experience in the Metaverse, enabling users to interact and work together in virtual worlds.

By fusing the virtual world with NLP, computer vision, and neural interface, AI also makes it possible for the Metaverse to implement several different purposes. As a result, AI is crucial to the Metaverse since it provides dependability and improves performance for a better experience. However, it is still unclear from a technology and application standpoint how AI might affect and contribute to the Metaverse.

We present a thorough analysis of the current AI-based works from the technical and application viewpoints; we then talk about their prospective applications in the Metaverse. The key contributions of our study are essentially summed up as follows. We examine cutting-edge AI-powered methodologies reliant on numerous technical facets, which exhibit tremendous promise for the establishment and growth of a Metaverse platform. For certain applications

that are more likely to be used in the Metaverse, a thorough study of the current AI-aided methodologies is offered. We discuss several intriguing Metaverse cases where AI has been used to improve the immersive experience and create user-focused services. Additionally, some potential AI research avenues for the Metaverse are discussed.

7 Concluding remarks

As the Metaverse becomes more prevalent and VR experiences become more advanced, the integration of AI with technologies such as IoT, blockchain, VR, AR, MR, and XR will continue to play an increasingly important role in shaping its development. This survey paper has explored the current state and future possibilities of AI in the Metaverse, including the potential benefits and challenges that come with this integration.

This survey offers a thorough analysis of how AI contributes to the creation of immersive Metaverse experiences. The goal of AI research is to build machines that comprehend natural language, interpret data, and take action on that data in the same way that humans do. AI can process a lot of data quicker and more effectively than humans. The Metaverse uses a variety of cutting-edge technologies in which AI plays an important role. We went over the key ideas behind AI as they pertain to learning paradigms in the different technologies of computer vision, XR, NLP, and blockchain. IoT and networking. We discussed the fundamental elements of the Metaverse (i.e., immersion creation, user interaction, NLP. The aforementioned technologies' principles are offered to give readers a better knowledge of their technical components and state-of-the-art, paving the way for a fully immersive Metaverse. To provide a future perspective of how AI technology might aid in the building of an XR environment that is as realistic as possible, we presented a thorough overview of the state-of-the-art in AI and examined the role of AI in achieving the Metaverse. The virtual environment will be a fantastic tool for people and those who inhabit it with this technology. By understanding these possibilities, we can better prepare for the exciting new era of VR and ensure that we are using AI ethically and responsibly. Ultimately, AI has the potential to greatly enhance the user experience in the Metaverse and improve the efficiency of virtual operations. As such, it is important to continue to explore and develop these technologies and integrate them in a way that maximizes their potential for positive impact.

Author contributions Dr. Mona (A) and Dr. Eman (B) wrote the manuscript textDr. Ashraf (C) and Dr. Aboul Ella (D) design the structure of the paperDr. Aboul Ella (D) gives the main ideas of the paper sections, revise all manuscript, revise and update figuresDr. Mona (A) and Dr. Ashraf (C) prepared all figuresAll authors reviewed the manuscript

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). We wish to confirm that there are no known conflicts of interest regarding this publication.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the

article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agarwal T, Arora H, Schneider J (2021) Learning urban driving policies using deep reinforcement learning. *IEEE International Intelligent Transportation Systems Conference (ITSC)*, Indianapolis, pp 607–614. <https://doi.org/10.1109/ITSC48978.2021.9564412>
- Ahmed E, Hassanien A, Sabry M, Alaa M (2023) Avatar videos. mendeley data, V1 <https://doi.org/10.17632/ygx3s9vmd9.1>
- Aijaz A, Sooriyabandara M (2018) The tactile internet for industries: a review. *Proc IEEE* 107(2):414–435
- Alex Yu, Sara Fridovich-Keil, Matthew Tancik, Qinhong Chen, Benjamin R, Angjoo K (2021) Plenoxels: radiance fields without neural networks. Preprint at <https://arxiv.org/abs/2112.05131>
- Alexey B, Chien-Yao W, Hong-Yuan ML (2020) Yolov4: Optimal speed and accuracy of object detection. Preprint at [arXiv:2004.10934](https://arxiv.org/abs/2004.10934)
- Alrubei SM, Ball EA, Rigelsford JM, Willis CA (2020) Latency and performance analyses of real-world wireless IoT-blockchain application. *IEEE Sens J* 20(13):7372–7383
- Amin G, Hossein K, Saeid S (2020) DeepFaceAR: deep face recognition and displaying personal information via augmented reality. In *2020 International Conference on Machine Vision and Image Processing (MVIP)*, IEEE, pp. 1–7, <https://doi.org/10.1109/MVIP49855.2020.9116873>
- Amirreza B, Igor D, Zachary OT (2017) Understanding dangerous play: a grounded theory analysis of high-performance drone racing crashes. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, New York, pp 485–496
- Baevski A, Hsu W-N, Xu Q, Babu A, Gu J, Auli M (2022) Data2vec: a general framework for self-supervised learning in speech, vision and language. Preprint at <https://arxiv.org/abs/2202.03555>
- Beer S (2015) Virtual museums: an innovative kind of museum survey. In: *Proceedings of the 2015 Virtual Reality International Conference*. pp 1–6
- Ben H (2023) The difference between ChatGPT and GPT-3. <https://dev.to/ben/the-difference-between-chatgpt-and-gpt-3-19dh>. Accessed 4 Feb 2023
- Ben M, Pratul PS, Matthew T, Jonathan TB, Ravi R, Ren N (2020) NeRF: representing scenes as neural radiance fields for view synthesis. Preprint at <https://doi.org/10.48550/arxiv.2003.08934>
- Bernardo M, Paulo D, João A, Beatriz SS (2019) Adaptive augmented reality user interfaces using face recognition for smart home control. In *International Conference on Human Systems Engineering and Design: Future Trends and Applications*, Springer, pp 15–19, https://doi.org/10.1007/978-3-030-27928-8_3
- Booth J, Roussos A, Zafeiriou S, Ponniah A, Dunaway D (2016) A 3d morphable model learnt from 10,000 faces. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp 5543–5555
- Menglei Chai, Tianjia Shao, Hongzhi Wu, Yanlin Weng, and Kun Zhou (2016) Autohair: fully automatic hair modeling from a single image. *ACM Trans Graph* 35(4)
- Chen D, Wang D, Zhu Y, Han Z (2021) Digital twin for federated analytics using a Bayesian approach. *IEEE Internet Things J* 8(22):16301–16312
- Chen Xieling, Xie Haoran, Tao Xiaohui (2022) Vision, status, and research topics of natural language processing. *Nat Lang Proc J* 1:100001. <https://doi.org/10.1016/j.nlp.2022.100001>. (ISSN 2949-7191)
- Chen L, Fu Q, Mu Y, Zeng L, Rezaeibagha F, Hwang M-S (2022) Blockchain-based random auditor committee for integrity verification. *Future Gener Comput Syst* 131:183–193
- Cheng S, Zhang Y, Li X, Yang L, Yuan X, Li SZ (2022) Roadmap toward the Metaverse: An AI perspective. *Innovation (Camb)* 3(5):100293. <https://doi.org/10.1016/j.xinn.2022.100293>. (PMID: 36032197; PMID: PMC9400115)
- Chowdhery A et al. (2022) PaLM: scaling language modeling with pathways. Preprint at <https://arxiv.org/abs/2204.02311>
- Christian R, Songyou P, Yiyi L, Andreas G (2021) KiloNeRF: speeding up neural radiance fields with thousands of tiny MLPs. Preprint at <https://arxiv.org/abs/2103.13744>

- Costa-jussà, et al. (2022) No language left behind: scaling human-centered machine translation. Preprint at <https://arxiv.org/abs/2207.04672>
- Datagen (2023) Guide: neural radiance field (NeRF). <https://datagen.tech/guides/synthetic-data/neural-radiance-field-nerf/>. Accessed 5 Feb 2023
- David F (2019) Generative Deep Learning. O'Reilly Media, Inc., ISBN: 9781492041948, <https://www.oreil.ly.com/library/view/generative-deep-learning/9781492041931/>
- Davis A, Murphy JD, Owens D, Khazanchi D, Zigurs I (2009) Avatars, people, and virtual worlds: foundations for research in Metaverses. *J Assoc Inf Syst* 10(2):90
- de la Fuente Julián, Prieto Pilar Lacasa, Martínez-Borda Rut (2022) Approaching Metaverses: mixed reality interfaces in youth media platforms. *New Techno Humanit* 2(2):136–145. <https://doi.org/10.1016/j.techum.2022.04.004>. (ISSN 2664-3294)
- Deng T, Zhang K, Shen Z-JM (2021) A systematic review of a digital twin city: a new pattern of urban governance toward smart cities. *J Manag Sci Eng* 6:125–134
- Deveci Muhammet, Gokasar Ilgin, Castillo Oscar, Daim Tugrul (2022) Evaluation of Metaverse integration of freight fluidity measurement alternatives using fuzzy Dombi EDAS model. *Comput Ind Eng*. <https://doi.org/10.1016/j.cie.2022.108773>
- NVIDIA Developers (2023) Megatron-turing natural language generation. <https://developer.nvidia.com/megatron-turing-natural-language-generation>. Accessed 4 Feb 2023
- Dirk Reiners, Reza Davahli Mohammad, Waldemar Karwowski, Carolina Cruz-Neira (2021) The combination of artificial intelligence and extended reality: a systematic review. *Front Virtual Reality*. <https://doi.org/10.3389/frvir.2021.721933>. (ISSN=2673-4192)
- Duan H, Li J, Fan S, Lin Z, Wu X, Cai W (2021) Metaverse for social good: a university campus prototype. In: Proceedings of the 29th ACM International Conference on Multimedia. Association for Computing Machinery, New York, pp 153–161
- Farsi M, Daneshkhah A, Hosseinian-Far A, Jahankhani H (2020) Digital twin technologies and smart cities. Springer, Cham. <https://doi.org/10.1007/978-3-030-18732-3>
- Ford TJ, Buchanan DM, Azeez A, Benrimoh DA, Kaloiani I, Bandeira ID, Hunegnaw S, Lan L, Gholmieh M, Buch V, Williams NR (2023) Taking modern psychiatry into the Metaverse: Integrating augmented, virtual, and mixed reality technologies into psychiatric care. *Front Digit Health* 24:5. <https://doi.org/10.3389/fgdth.2023.1146806>. (PMID: 37035477; PMCID: PMC10080019)
- Gadekallu T, Huynh-The T, Wang W, Yenduri G, Ranaweera P, Pham V, Costa DB, Liyanage M (2022). Blockchain for the Metaverse: a review
- Gao Y, Wu W, Si P, Yang Z, Yu FR (2021) B-rest: blockchain enabled resource sharing and transactions in fog computing. *IEEE Wirel Commun* 28(2):172–180
- Gema B-O, David C (2013) Comparative study of text clustering techniques in virtual worlds. In Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics, WIMS '13, New York, pp 1–8
- Gema BO, Maria DR-M, David C, David FB (2012) Clustering avatars behaviours from virtual worlds interactions. In Proceedings of the 4th International Workshop on Web Intelligence & Communities. WI&C '12, New York, pp 1–7
- Genay ACS, Lecuyer A, Hachet M (2021) Being an avatar for real: a survey on virtual embodiment in augmented reality. *IEEE Trans Vis Comput Graphics*
- Ghosh A, Chakraborty D, Law A (2018) Artificial intelligence in internet of things. *CAAI Trans Intell Technol* 3(4):208–218. <https://doi.org/10.1049/trit.2018.1008>
- Gokasar Ilgin, Pamucar Dragan, Deveci Muhammet, Gupta Brij B, Martinez Luis, Castillo Oscar (2023) Metaverse integration alternatives of connected autonomous vehicles with self-powered sensors using fuzzy decision making model. *Information Sciences* 642:119192. <https://doi.org/10.1016/j.ins.2023.119192>. (ISSN 0020-0255)
- Güven İ, Ballı O (2022) Empowering Metaverse through artificial intelligence
- Hassan MU, Rehmani MH, Chen J (2019) Privacy preservation in blockchain based IoT systems: integration issues, prospects, challenges, and future research directions. *Future Gener Comput Syst* 97:512–529
- Haythem B, David K, Jan K (2019) Accurate object detection system on hololens using yolo algorithm. In 2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO), IEEE, pp 219–224
- Hbali Youssef, Ballihi Lahoucine, Sadgal Mohammed, El Fazziki Abdelaziz (2016) Face detection for augmented reality application using boosting-based techniques. *Int J Interact Multim Artif Intell* 4(2):22–28
- Hongyu L, Tianqi H (2019) Towards diverse anime face generation: active label completion and style feature network. In Eurographics (Short Papers), pp 65–68

- Hu F, Xia G-S, Wang Z, Huang X, Zhang L, Sun H (2015) Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. <https://doi.org/10.1109/JSTARS.2015.2444405>
- Huanle Z, Bo H, Cheuk YI, Prasant M (2020) limmer: Accelerating 3d semantic segmentation for mobile augmented reality. In 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), IEEE, pp 603–612
- Hujon AV, Singh TD, Amitab K (2023) Transfer learning based neural machine translation of english-Khasi on low-resource settings. *Proc Comput Sci* 218:1–8. <https://doi.org/10.1016/j.procs.2022.12.396>
- Huynh-The T, Pham Q-V, Pham X-Q, Nguyen TT, Han Z, Kim D-S (2022) Artificial intelligence for the Metaverse: a survey. Preprint at <https://arxiv.org/abs/2202.10336>
- Igor C, Petra U, Vojtech J, Zuzana F, Marie J, Jiri S, Barbora K (2017) Generating various composite human faces from real 3d facial images. *Visual Comput* 33(4):443–458
- Isha S (2022) NVIDIA research turns 2D photos into 3D scenes in the blink of an AI. nvidia. <https://blogs.nvidia.com/blog/2022/03/25/instant-nerf-research-3d-ai/>. Accessed 6 Nov 2022
- Jakub L, Vladimir B (2019) GANs in Action, Manning ISBN 9781617295560. <https://www.manning.com/books/gans-in-action>
- Jan S, Jonatan A (2018) Object detection in augmented reality. Master's theses in mathematical sciences. <https://lup.lub.lu.se/student-papers/search/publication/8964243>
- Jankowski J, Hachet M (2013) A survey of interaction techniques for interactive 3d environments, in Eurographics 2013-STAR
- Jenny L, Xingwen G, Jingyu S, Chenfanfu J, Yixin Zu, Song-Chun Z (2016) A virtual reality platform for dynamic human-scene interaction. In SIGGRAPH ASIA 2016 virtual reality meets physical reality: Modelling and simulating virtual humans and environments, pp 1–4
- Jing L, Tian Y (2021) Self-supervised visual feature learning with deep neural networks: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43(11): 4037–4058. <https://doi.org/10.48550/arXiv.1902.06162>
- Jonathan S (2022) Getting started with NVIDIA instant NeRFs. nvidia. <https://developer.nvidia.com/blog/getting-started-with-nvidia-instant-nerfs/>. Accessed 6 Nov 2022
- Jonathan TB, Ben M, Matthew T, Peter H, Ricardo M-B, Pratul PS (2021) Mip-NeRF: a multiscale representation for anti-aliasing neural radiance fields. Preprint at <https://arxiv.org/abs/2103.13415>
- Joseph R, Ali F (2018) Yolov3: An incremental improvement. Preprint at <https://arxiv.org/abs/1804.02767>
- Kanter TG (2021) The Metaverse and extended reality with distributed IoT. *IEEE Internet of Things Magazine (IoT)*
- Khan Wazir Zada, Ahmed Ejaz, Hakak Saqib, Yaqoob Ibrar, Ahmed Arif (2019) Edge computing: a survey. *Futur Gener Comput Syst* 97:219–235. <https://doi.org/10.1016/j.future.2019.02.050>
- Kido Daiki, Fukuda Tomohiro, Yabuki Nobuyoshi (2021) Assessing future landscapes using enhanced mixed reality with semantic segmentation by deep learning. *Adv Eng Inform* 48:101281
- Kim K, Park S (2023) AOBERT: all-modalities-in-one BERT for multimodal sentiment analysis. *Inform Fusion* 92:37–45. <https://doi.org/10.1016/j.inffus.2022.11.022>
- Kiumarsi B, Vamvoudakis KG, Modares H, Lewis FL (2018) Optimal and autonomous control using reinforcement learning: a survey. *IEEE Transactions on Neural Networks and Learning Systems* 29(6):2042–2062. <https://doi.org/10.1109/TNNLS.2017.2773458>
- Ko Tae-young, Lee Seung-ho (2020) Novel method of semantic segmentation applicable to augmented reality. *Sensors* 20(6):1737
- Kocur M, Graf S, Schwind V (2020) The impact of missing fingers in virtual reality. In: 26th ACM Symposium on Virtual Reality Software and Technology. In: VRST '20, Association for Computing Machinery ACM ISBN 978-1-4503-7619-8/20/11 <https://doi.org/10.1145/3385956.3418973>.
- Koichi H, Kentaro T, Tianqi L, Hiroto H, Yusuke U (2018) Full-body high-resolution anime generation with progressive structure-conditional generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops
- Kotsiantis SB, Zaharakis I, Pintelas P et al (2007) Supervised machine learning: a review of classification techniques. *Emerg Artif Intell Appl Comput Eng* 160(1):3–24
- Kuzlu M, Fair C, Guler O (2022) Role of artificial intelligence in the internet of things (IoT) cybersecurity. *Discov Internet Things* 1(1):2021. <https://doi.org/10.1007/s43926-020-00001-4>
- Lalapura VS, Amudha J, Sathesh HS (2022) Recurrent neural networks for edge intelligence: a survey. *ACM Comput Surv (CSUR)* 54(4):1–38. <https://doi.org/10.1145/3448974>
- Lambert Nick (2021) Beyond nfts: a possible future for digital art. *ITNOW* 63(3):8–10

- Lee L-H, Braud T, Zhou P, Wang L, Xu D, Lin Z, Kumar A, Bermejo C, Hui P (2021) All one needs to know about Metaverse: a complete survey on technological singularity, virtual ecosystem, and research agenda. Preprint at <https://arxiv.org/abs/2110.05352>
- Leippold M (2022) Thus spoke GPT-3: interviewing a large-language model on climate finance. *Financ Res Lett*. <https://doi.org/10.1016/j.frl.2022.103617>
- Leotta MJ, Long C, Jacquet B, Zins M, Lipsa D, Shan J, Xu B, Li Z, Zhang X, Chang S-F, et al. (2019) Urban semantic 3D reconstruction from multiview satellite imagery. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*
- Li Yonghai, Jinran Yu, Wei Yichen, Wang Yifei, Feng Zhenyu, Cheng Liuqi, Huo Ziwei, Lei Yanqiang, Sun Qijun (2023) Recent progress in self-powered wireless sensors and systems based on TENG. *Sensors* 23(3):1329. <https://doi.org/10.3390/s23031329>
- Liang-Chieh C, Yukun Z, George P, Florian S, Hartwig A (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp 801–818. <https://doi.org/10.48550/arXiv.1802.02611>
- Lin W, Yujeong C, Kuk-Jin Y (2021) Dual transfer learning for event-based end-task prediction via pluggable event-to-image translation. *ICCV*. <https://doi.org/10.48550/arXiv.2109.01801>
- Lin W, Yujeong C, Sung-Hoon Y, Tae-Kyun K, Kuk-Jin Y (2021) Evidistill: Asynchronous events to end-task learning via bidirectional reconstruction-guided cross-modal knowledge distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp 608–619
- Lingjie L, Jiatao G, Kyaw ZL, Tat-Seng C, Christian T (2020) Neural sparse voxel fields. Preprint at <https://arxiv.org/abs/2007.11571>
- Liu L, Habermann M, Rudnev V, Sarkar K, Gu J, Theobalt C (2021) Neural actor: neural free-view synthesis of human actors with pose control. Preprint at <https://arxiv.org/abs/2106.02019>
- Lu E, Miller J, Pereira N, Rowe A (2021) FLASH: Videoembeddable ar anchors for live events. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, pp 489–497
- Luo Y, Jin H, Li P (2019) A blockchain future for secure clinical data sharing: a position paper. In *Proceedings of the ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization*, pp 23–27
- Luo H, Nagano K, Kung H-W, Xu Q, Wang Z, Wei L, Hu L, Li H (2021) Normalized avatar synthesis using stylegan and perceptual refinement. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp 11662–11672
- Luyang L, Hongyu L, Marco G (2019) Edge assisted realtime object detection for mobile augmented reality. In *The 25th Annual International Conference on Mobile Computing and Networking*, pp 1–16
- Ma X, Kang K, Zhu R, Wu H, Zhou K (2021) Free-form scanning of non-planar appearance with neural trace photography. Vol 40. No 4. Association for Computing Machinery, New York
- Martin S, Stefan M, Karl A, Horst-Michael G (2018) Complex-yolo: an euler-region-proposal for real-time 3d object detection on point clouds. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp 0–0
- Mary KY, John JR, Bobby B (2015) Dyadic interactions with avatars in immersive virtual environments: high fiving. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception SAP '15*, New York, pp 119–126
- Matthieu N, Laura A, Flavio Di G, Mauro M, Luca MA, Andrea B (2021) Mapping the nft revolution: market trends, trade networks and visual features. Preprint at [arXiv:2106.00647](https://arxiv.org/abs/2106.00647)
- Menandro R, Tomoki H, Taiki F, Yasuhide O, Takeshi O (2018) Occlusion handling using semantic segmentation and visibility-based rendering for mixed reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp 1–8
- Meta (2021) Introducing horizon workrooms: remote collaboration reimagined. <https://about.fb.com/news/2021/08/introducing-horizon-workrooms-remote-collaboration-reimagined/>. Accessed Aug 2021
- Minerva R, Lee GM, Crespi N (2020) Digital twin in the IoT context: a survey on technical features, scenarios, and architectural models. *Proc IEEE* 108(10):1785–1824
- Moradi Ashkan, Shekofteh Yasser (2023) Spoken language identification using a genetic-based fusion approach to combine acoustic and universal phonetic results. *Comput Elect Eng* 105:108549. <https://doi.org/10.1016/j.compeleceng.2022.108549>. (ISSN 0045-7906)
- Muhammed D, Dragan P, Ilgin G, Mario K, Brij BG, Tugrul D (2023) Evaluation of Metaverse traffic safety implementations using fuzzy Einstein based logarithmic methodology of additive weights and TOPSIS method. *Technol Forecast Soc Change* 194:122681. <https://doi.org/10.1016/j.techfore.2023.122681>. (ISSN 0040-1625)

- Murphy D (2017) Building a hybrid virtual agent for testing user empathy and arousal in response to avatar (micro-)expressions. In: Proceedings of the 23rd ACM Symposium on virtual reality software and technology. In: VRST '17, Association for Computing Machinery, New York
- Nahuel AM, Jorge SI, María JA (2020) Scalable integration of image and face-based augmented reality. In International Conference on Augmented Reality, Virtual Reality and Computer Graphics, Springer, pp 232–242 https://doi.org/10.1007/978-3-030-58465-8_18
- Nahuel AM, Jorge SI, María JA. (2020) Scalable integration of image and face based augmented reality. In International Conference on Augmented Reality, Virtual Reality and Computer Graphics, Springer, pp 232–242
- Navarro F, Fdez J, Garzón M, Roldán JJ, Barrientos A (2017) Integrating 3D reconstruction and virtual reality: a new approach for immersive teleoperation. In: ROBOT
- Nextech AR (2022) What are the different types of augmented reality? <https://www.nextechar.com/blog/what-are-the-different-types-of-augmented-reality>. Accessed 12 Dec 2022
- Ning H, Wang H, Lin Y, Wang W, Dhelim S, Farha F, Ding J, Daneshmand M (2021) A survey on Metaverse: the state-of-the-art, technologies, applications, and challenges. Preprint at [arXiv:2111.09673](https://arxiv.org/abs/2111.09673)
- NLLB Team (2022) <https://github.com/facebookresearch/fairseq/tree/nllb>. Accessed 2 Apr 2023
- OpenAI (2022) Dalle-2. <https://openai.com/dall-e-2/>. Accessed 3 Dec 2022
- OpenAI (2023) ChatGPT: Optimizing language models for dialogue, <https://openai.com/blog/chatgpt/>. Accessed 8 Aug 2023
- Pamucar Dragan, Devci Muhammet, Gokasar Ilgin, Delen Dursun, Köppen Mario, Pedrycz Witold (2023) Evaluation of Metaverse integration alternatives of sharing economy in transportation using fuzzy Schweizer-Sklar based ordinal priority approach. *Decis Support Syst* 171:113944. <https://doi.org/10.1016/j.dss.2023.113944>. (ISSN 0167-9236)
- Park S-M, Kim Y-G (2022) A Metaverse: taxonomy, components, applications, and open challenges. *IEEE Access* 10:4209–4251. <https://doi.org/10.1109/ACCESS.2021.3140175>
- Peer S, Max S, Sven B (2019) Semantic interaction in augmented reality environments for Microsoft HoloLens. In 2019 European Conference on Mobile Robots (ECMR), IEEE, pp 1–6
- Pereira N, Rowe A, Farb MW, Liang I, Lu E, Riebling E (2021) Arena: The augmented reality edge networking architecture, in IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE 2021:479–488
- Promwongsa N, Ebrahimzadeh A, Naboulsi D, Kianpisheh S, Belqasmi F, Glioth R, Crespi N, Alfandi O (2020) A comprehensive survey of the tactile internet: state-of-the-art and research directions. *IEEE Commun Surv Tutor* 23(1):472–523
- Qamar Ara, Anwar Zahid, Afzal Mehreen (2023) A systematic threat analysis and defense strategies for the Metaverse and extended reality systems. *Comput Secur* 128:103127. <https://doi.org/10.1016/j.cose.2023.103127>. (ISSN 0167-4048)
- Qi CR, Su H, Mo K, Guibas LJ (2016) PointNet: deep learning on point sets for 3D classification and segmentation. Preprint at <https://arxiv.org/abs/1612.00593>
- Qi CR, Liu W, Wu C, Su H, Guibas LJ (2017) Frustum PointNets for 3D object detection from RGB-D data. Preprint at <https://arxiv.org/abs/1711.08488>
- Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M (2022) Hierarchical text-conditional image generation with CLIP latents. Preprint at <https://arxiv.org/abs/2204.06125>
- Rathore MM, Shah SA, Shukla D, Bentafat E, Bakiras S (2021) The role of AI, machine learning, and big data in digital twinning: a systematic literature review, challenges, and opportunities. *IEEE Access* 9:32030–32052
- Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H (2016) Generative adversarial text to image synthesis. Preprint at <https://arxiv.org/abs/1605.05396>
- Ren Shaoqing, He Kaiming, Girshick Ross, Sun Jian (2015) Faster rcnn: towards real-time object detection with region proposal networks. *Adv Neural Inform Process Syst* 28:91–99
- Romero J, Tzionas D, Black MJ (2017) Embodied hands: modeling and capturing hands and bodies together. *ACM Trans Graph*. 36(6):1–17
- Sagar M (2018) Integrating yolo object detection with augmented reality for ios apps. In 2018 9th IEEE Annual Ubiquitous Computing, Journal of Latex Class Files, vol. 14, no. 8, September 2021 57, Electronics & Mobile Communication Conference (UEMCON), IEEE, pp 585–589
- Salah Mohammed, Halbusi Hussam AI, Abdelfattah Fadi (2023) May the force of text data analysis be with you: Unleashing the power of generative AI for social psychology research. *Comput Human Behav: Artif Humans* 1(2):100006. <https://doi.org/10.1016/j.chbah.2023.100006>. (ISSN 2949-8821)
- Seong H, Hyun J, Kim E (2020) F0snet: an end-to-end trainable deep neural network for scene recognition. *IEEE Access* 8:82 066-82 077

- Shi Feifei, Ning Huansheng, Zhang Xiaohong, Li Rongyang, Tian Qiaohui, Zhang Shiming, Zheng Yuanyuan, Guo Yudong, Daneshmand Mahmoud (2023) A new technology perspective of the Metaverse: its essence, framework and challenges. *Digit Commun Netw*. <https://doi.org/10.1016/j.dcan.2023.02.017>. (ISSN 2352-8648)
- Sodhro AH, Pirbhulal S, Sangaiah AK (2018) Convergence of IoT and product lifecycle management in medical health care. *Future Gener Comput Syst* 86:380–391
- Steinbach E, Strese M, Eid M, Liu X, Bhardwaj A, Liu Q, Al-Ja'afreh M, Mahmoodi T, Hassen R, El Sadiq A et al (2018) Haptic codecs for the tactile internet. *Proc IEEE* 107(2):447–470
- Stojanovic N, Milenovic D (2018) Data-driven digital twin approach for process optimization: an industry use case. In *IEEE International Conference on Big Data (Big Data)*. IEEE, pp 4202–4211
- Stotko P, Krumpen S, Hullin MB, Weinmann M, Klein R (2019) SLAMCast: Large-scale, real-time 3D reconstruction and streaming for immersive multi-client live telepresence. *IEEE Trans Vis Comput Graphics* 25(5):2102–2112
- Sun Mengting, Xie Linshan, Liu Yang, Li Kecheng, Jiang Bin, Yeting Lu, Yang Yuping, Hai Yu, Song Yuanlin, Bai Chunxue, Yang Dawei (2022) The Metaverse in current digital medicine. *Clin eHealth* 5:52–57. <https://doi.org/10.1016/j.ceh.2022.07.002>. (ISSN 2588-9141)
- Takayuki N, Takashi K (2019) Semi-automatic creation of an anime-like 3d face model from a single illustration. In *2019 International Conference on Cyberworlds (CW)*, IEEE, pp 53–56
- Talasila V, Narasingarao MR, MuraliMohan V (2022) Optimized GAN for text-to-image synthesis: hybrid whale optimization algorithm and dragonfly algorithm. *Adv Eng Softw* 173:103–222. <https://doi.org/10.1016/j.advengsoft.2022.103222>
- Tan YX, Lee CP, Neo M, Lim KM (2022) Text-to-image synthesis with self-supervised learning. *Pattern Recogn Lett* 157:119–126. <https://doi.org/10.1016/j.patrec.2022.04.010>
- Tanja K, Danish A, Robert G, Sebastian M, Jan-Niklas V-A (2020) The user experience of reading in virtual reality—finding values for text distance, size, and contrast. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, IEEE, pp 1–6 <https://doi.org/10.48550/arXiv.2004.01545>
- Tanzi L, Piazzolla P, Porpiglia F et al (2021) Real-time deep learning semantic segmentation during intra-operative surgery for 3D augmented reality assistance. *Int J CARS* 16:1435–1445. <https://doi.org/10.1007/s11548-021-02432-y>
- Tao F, Zhang H, Liu A, Nee AYC (2019) Digital twin in industry: state-of-the-art. *IEEE Trans Ind Inform* 15(4):2405–2415
- Tianyang S, Yi Y, Changjie F, Zhengxia Z, Zhenwei S, Yong L (2019) Face-to-parameter translation for game character auto-creation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp 161–170
- Vázquez Juan José, Arjona Jamie, Linares M^aPaz, Casanovas-Garcia Josep (2020) A comparison of deep learning methods for urban traffic forecasting using floating car data. *Transp Res Proc* 47:195–202. <https://doi.org/10.1016/j.trpro.2020.03.079>
- Venkatesan Mythreye, Mohan Harini, Ryan Justin R, Schürch Christian M, Nolan Garry P, Frakes David H, Coskun Ahmet F (2021) Virtual and augmented reality for biomedical applications. *Cell Rep Med* 2(7):100348. <https://doi.org/10.1016/j.xcrm.2021.100348>. (ISSN 2666-3791)
- Vernaza A, Armuelles VI, Ruiz I (2012) Towards to an open and interoperable virtual learning environment using Metaverse at University of Panama. In *2012 Technologies Applied to Electronics Teaching (TAEE)*, pp 320–325
- Vondráček Artin, Baggili Ibrahim, Casey Peter, Mekni Mehdi (2023) Rise of the Metaverse's immersive virtual reality malware and the man-in-the-room attack & defenses. *Comput Secur* 127:102923. <https://doi.org/10.1016/j.cose.2022.102923>. (ISSN 0167-4048)
- Wang T, Zhang B, Zhang T, Gu S, Bao J, Baltrušaitis T, Shen J, Chen D, Wen F, Chen Q, Guo B (2022) Rodin: a generative model for sculpting 3D digital avatars using diffusion. Preprint at <https://arxiv.org/abs/2212.06135>
- Warpefelt H, Verhagen H (2015) Towards an updated typology of non-player character roles. In: *Proceedings of the International Conference on Game and Entertainment Technologies*. pp 1–9
- Wei X, Yin L, Zhu Z, Ji Q (2004) Avatar-mediated face tracking and lip reading for human computer interaction. In: *Proceedings of the 12th Annual ACM International Conference on Multimedia*. In: MULTIMEDIA '04, Association for Computing Machinery, New York
- Wei L, Dragomir A, Dumitru E, Christian S, Scott R, Cheng-Yang F, Alexander CB (2016) Ssd: Single shot multibox detector. In *European conference on computer vision*, Springer, pp 21–37
- Wibowo S, Sandikapura T (2019) Improving data security, interoperability, and veracity using blockchain for one data governance, case study of local tax big data. *Int Conf ICT Smart Soc (ICISS)* 7:1–6

- Wu H, Han H, Wang X, Sun S (2020) Research on artificial intelligence enhancing internet of things security: a survey. *IEEE Access* 8:153826–153848. <https://doi.org/10.1109/ACCESS.2020.3018170>
- Wu M, Wang Y, Hu Q, Yu J (2020) Multi-view neural human rendering. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp 1682–1691
- Xiao A, Huang J, Guan D, Cui K, Lu S, Shao L (2022) PolarMix: a general data augmentation technique for LiDAR point clouds. Preprint at <https://arxiv.org/abs/2208.00223>
- Xie J, Yu FR, Huang T, Xie R, Liu J, Liu Y (2019) A survey on the scalability of blockchain systems. *IEEE Network* 33(5):166–173
- Xu X, Sun G, Luo L, Cao H, Yu H, Vasilakos AV (2021) Latency performance modeling and analysis for hyperledger fabric blockchain network. *Inform Proc Manag* 58(1):102436
- Xueshi L, Difeng Y, Hai-Ning L, Wenge X, Yuzheng C, Xiang L, Khalad H (2020) Exploration of hands-free text entry techniques for virtual reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, pp 344–349 <https://doi.org/10.1109/ISMAR50242.2020.00061>
- Yakun H, Xiuquan Q, Haowen W, Xiang S, Schahram D, Ping Z (2022) Multi-player immersive communications and interactions in Metaverse: challenges, architecture, and future directions. Preprint at <https://arxiv.org/abs/2210.06802>
- Yan Y, Mao Y, Li B (2018) Second: sparsely embedded convolutional detection. *Sensors* 18(10):3337. <https://doi.org/10.3390/s18103337>
- Yang B, Luo W, Urtasun R (2019) PIXOR: real-time 3D object detection from point clouds. Preprint at <https://arxiv.org/abs/1902.06326>
- Yang Q, Yetong Z, Huang H, Xiong Z, Kang J, Zheng Z (2022) Fusing blockchain and AI with Metaverse: a survey. Preprint at arXiv.2201.03201
- Yanghua J, Jiakai Z, Minjun L, Yingtao T, Huachun Z, Zhihao F (2017) Towards the automatic anime characters creation with generative adversarial networks. Preprint at <https://arxiv.org/abs/1708.05509>
- Yassir El F, Salah-ddine K (2020) Augmented reality types and popular use cases. In *Proceedings of the 1st International Conference of Computer Science and Renewable Energies - ICCSRE*, SciTePress, pp 107–110
- Yiu N (2021) An overview of forks and coordination in blockchain development. Preprint at <https://arxiv.org/abs/2102.10006>
- Yogesh KD et al (2022) Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice, and policy. *Int J Inform Manag* 66:102542. <https://doi.org/10.1016/j.ijinfomgt.2022.102542>. (ISSN 0268-4012)
- Kiwoong Yoo, Roman Welden, Kelly Hewett, Michael Haenlein, The merchants of meta: A research agenda to understand the future of retailing in the Metaverse, *Journal of Retailing*, 2023, ISSN 0022-4359, <https://doi.org/10.1016/j.jretai.2023.02.002>.
- Yu Wu, Ma Lisha, Yuan Xiaofang, Li Qingnan (2023) Human-machine hybrid intelligence for the generation of car frontal forms. *Adv Eng Inform* 55:101906. <https://doi.org/10.1016/j.aei.2023.101906>. (ISSN 1474-0346)
- Yungang W, Xiaoran Q, Xiaoye T, Xiaohang Y, Bo S, Xiaoming Z (2015) The design of a visual tool for the quick customization of virtual characters in OSSL. In *2015 International Conference on Cyberworlds (CW)*, pp 314–320
- Zawish M, Dharejo F, Khowaja S, Dev K, Davy S, Qureshi NM F, Bellavista P (2022) AI and 6G into the Metaverse: fundamentals, challenges and future research trends. Preprint at arXiv.2208.10921
- Żelazarczyk M, Mańdziuk J (2023) Cross-modal text and visual generation: a systematic review. Part 1: image to text. *Inform Fusion* 93:302–329. <https://doi.org/10.1016/j.inffus.2023.01.008>
- Zhang Y, Zhang X, Nana D (2021) Research on 3D object detection method based on convolutional attention mechanism. *J Phys: Conf Ser*. <https://doi.org/10.1088/1742-6596/1848/1/012097>
- Zhao Yuheng, Jiang Jinjing, Chen Yi, Liu Richen, Yang Yalong, Xue Xiangyang, Chen Siming (2022) Metaverse: perspectives from graphics, interactions and visualization. *Visual Inform* 6(1):56–67. <https://doi.org/10.1016/j.visinf.2022.03.002>. (ISSN 2468-502X)
- Zhou Y, Tuzel O (2018) VoxelNet: end-to-end learning for point cloud based 3D object detection. *IEEE/CVF Conf Comput Vision Pattern Recog* 2018:4490–4499
- Zhu H (2022) MetaOnce: a Metaverse framework based on multi-scene relations and entity-relation-event game