Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Survey paper

Artificial intelligence for the metaverse: A survey

Thien Huynh-The [a,*], Quoc-Viet Pham [b], Xuan-Qui Pham [c], Thanh Thi Nguyen [d], Zhu Han [e], Dong-Seong Kim [c]

[a] *Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology and Education, Ho Chi Minh City, 70000, Viet Nam*
[b] *Korean Southeast Center for the 4th Industrial Revolution Leader Education, Pusan National University, Busan 46241, Republic of Korea*
[c] *ICT Convergence Research Center, Kumoh National Institute of Technology, Gyeongsangbuk 39177, Republic of Korea*
[d] *School of Information Technology, Deakin University, Waurn Ponds 3216, Australia*
[e] *Department of Electrical and Computer Engineering, University of Houston, Houston 77004, USA*

ARTICLE INFO

ABSTRACT

Along with the massive growth of the Internet from the 1990s until now, various innovative technologies have been created to bring users breathtaking experiences with more virtual interactions in cyberspace. Many virtual environments have been developed with immersive experience and digital transformation, but most are incoherent instead of being integrated into a platform. In this context, metaverse has been introduced as a shared virtual world that is fueled by many emerging technologies. Among such technologies, artificial intelligence (AI) has shown the great importance of enhancing immersive experience and enabling human-like intelligence of virtual agents. In this survey, we make a beneficial effort to explore the role of AI, including machine learning algorithms and deep learning architectures, in the foundation and development of the metaverse. As the main contributions, we convey a comprehensive investigation of AI-based methods concerning several technical aspects (e.g., natural language processing, machine vision, blockchain, networking, digital twin, and neural interface) that have potentials to build virtual worlds in the metaverse. Furthermore, several primary AI-aided applications, including healthcare, manufacturing, smart cities, and gaming, are studied to be promisingly deployed in the virtual worlds. Finally, we conclude the key contribution and open some future research directions of AI for the metaverse. Serving as a foundational survey, this work will help researchers, including experts and non-experts in related fields, in applying, developing, and optimizing AI techniques to polish the appearance of virtual worlds and improve the quality of applications built in the metaverse.

1. Introduction

Since Facebook rebranded itself as Meta, announced by Mark Zuckerberg in October 2021, the marvelous concept of the new name has become a hot trend on social media and received huge attention and much more discussions by various communities, including academia and industry. Besides Meta, some big tech companies have some metaverse investment and development activities, such as Microsoft bought Activision Blizzard, a video game holding company, for $68.7 billion as a deal of gaming expansion into the metaverse. Recently, Metaverse Group, a metaverse real estate investment company bought a parcel of land on a decentralized virtual reality platform known as Decentraland for a shocking price $2.43 million and recorded as the highest-ever amount for virtual real estate. A famous rapper who bought a plot of land in the Sandbox metaverse for $450,000 is Snoop Dogg, in which this rapper can hold virtual events like music festivals and concerts to bring an immersive experience to the audience participating in the

virtual world via the virtual reality technology. Soon, the metaverse is realized as the next big technology and currently attracting online game makers, internet finance businesses, social networks, and other technology leaders. The Seoul metropolitan government just very recently announced a plan called Metaverse Seoul that creates a virtual communication ecosystem for all municipal administrative areas, such as culture, tourism, economic, educational, and civic service. Besides providing different business support services and facilities, Metaverse Seoul will offer some specialized services for people with disabilities to take pleasure in safety and convenient content using extended reality (XR) technology. Based on the analysis of Bloomberg Intelligence, the global metaverse revenue opportunity will increase from USD 500 billion in 2020 to USD 800 billion in 2024, in which the online game industry will take half of the global revenue (Kanterman and Naidu, 2021). Remarkably, video game companies and studios have some plans to upgrade existing traditional games to three-dimensional (3D)
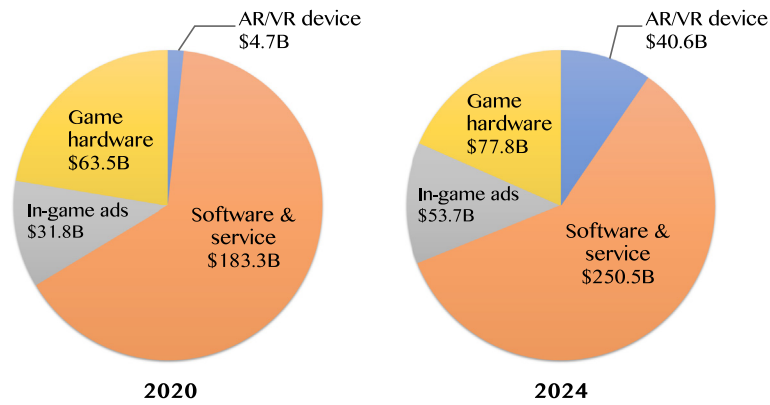

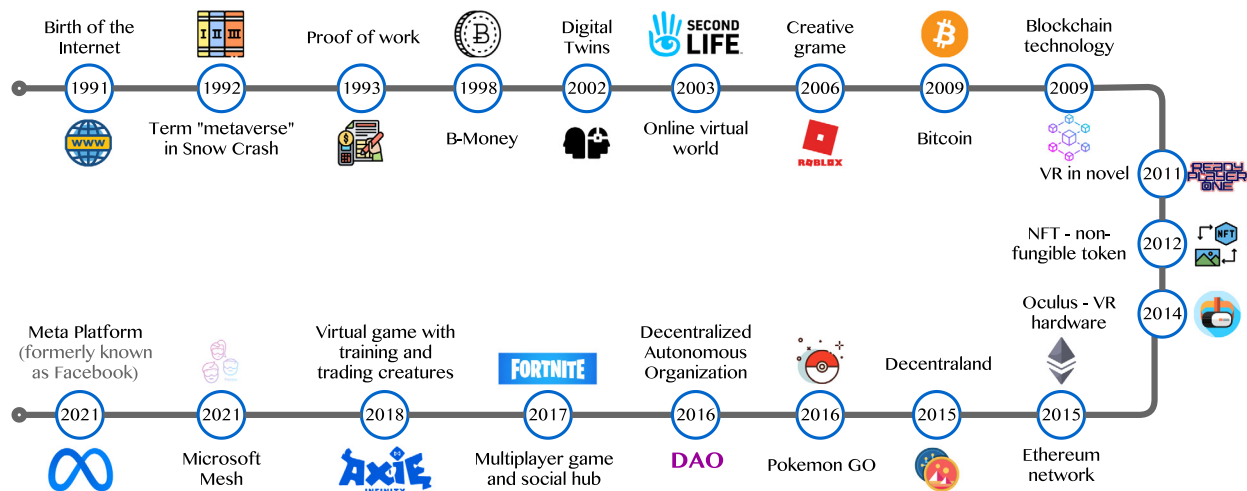Fig. 1. Gaming revenue growth aided by 3D virtual worlds.

Fig. 2. A timeline of the metaverse development involving primary events from 1991 to 2021.

virtual worlds convolving social networks, in which some attractive activities, such as live entertainment and media advertising events, can be held besides gaming. In Fig. 1, the revenue of virtual reality (VR) hardware and in-game advertisement significantly increases through the advancement of virtual activities in the metaverse.

The metaverse is not a new idea because it has circulated along with the development of the Internet and other technologies for decades. Fig. 2 describes the timeline of the metaverse development that involves many primary events, from the birth of the Internet and the first mention in literature to the first virtual world project with Second Life and recent metaverse projects of big tech companies like Microsoft and Facebook. Metaverse is the term formed by combining Meta and Universe, which may be first mentioned in the dystopian cyberpunk novel Snow Crash in 1992 to describe a virtual reality world called the matrix (Hackl, 2021). At present, the metaverse is defined as a shared virtual 3D world or even multiple cross-platform worlds that can provide users with a comprehensively immersive experience, including interactive and collaborative activities. Besides virtual places and constructions fixed in the virtual world, many other entities, such as objects, user identities, and digital goods, can be exchanged between different virtual worlds and even reflected into the reality world (Park and Kim, 2022). Few recent years have witnessed an unprecedented explosion of the metaverse, mostly derived from 3D gaming, which is fueled by the improvement of hardware (e.g., big data storage infrastructure, wireless communication networks, built-in sensors, and graphic processing unit—GPU) and the optimization of software (e.g., resource allocation in communications, language processing, and computer vision) to build the virtual world more solidly

and creatively. Different from the traditional metaverse modality that limits immersive experience poorly by insufficient data, the new one not only generates a huge new source of user and behavioral data for enterprises (where users freely make creative content) but also presents a plentiful foundation to deploy artificial intelligence (AI) into various domains, such as natural language processing, computer vision, and neural interface. Besides, a standard platform built for a modern metaverse should satisfy the following characteristics: virtual world, persistency, scalability, always-on with synchronicity, financial allowance, decentralization, security, and interoperability. In Radoff (2021), a metaverse platform can include several layers (see Fig. 3) which are expressed as follows:

- *Infrastructure*: 5G, 6G, WiFi, cloud, data center, central processing units, and GPUs.
- *Human interface*: mobile, smartwatch, smartglasses, wearable devices, head-mounted display, gestures, voice, and electrode bundle.
- *Decentralization*: edge computing, AI agents, blockchain, and microservices.
- *Spatial computing*: 3D engines, VR, augmented reality (AR), XR, geospatial mapping, and multitasking.
- *Creator economy*: design tools, asset markets, E-commerce, and workflow.
- *Discovery*: advertising networks, virtual stores, social curation, ratings, avatar, and chatbot.
- *Experience*: games, social, E-sports, shopping, festivals, events, learning, and working.
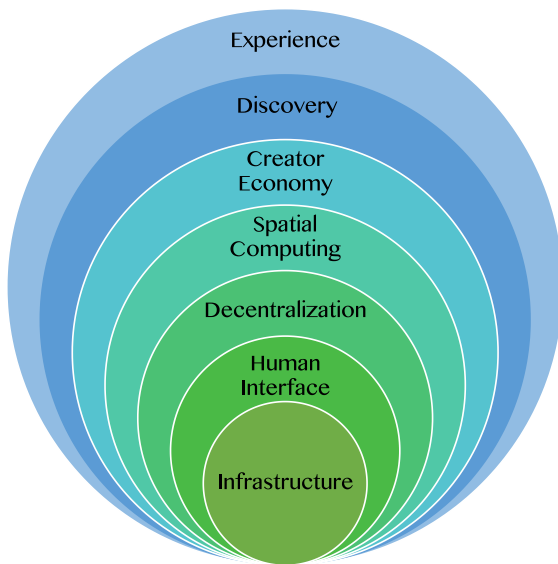
Fig. 3. Seven layers of a metaverse platform.

It is not hard to find out the presence of AI inside layers, with machine learning (ML) algorithms and deep learning (DL) architectures, along with their importance in many diversified aspects. For instance, many ML algorithms (Kotsiantis et al., 2007; Hu et al., 2015; Sheikhpour et al., 2017; Kiumarsi et al., 2018) for supervised, unsupervised, semi-supervised, and reinforcement learning were applied in classification and regression models for voice recognition and other language processing tasks that enable system agents to understand user commands. With the input data of sensor-based signals collected by multiple devices, such as mobiles, smartwatches, and other wearable devices, the complex patterns of human actions can be analyzed and learned for some applications like physical activity recognition that allows a system to perceive user activities and interactions in the virtual world. Recently, DL, which consisted of recurrent neural network (RNN), long short-term memory (LSTM), and convolutional neural network (CNN) architectures, has emerged as a powerful AI tool to deal with the practical issue of understanding complicated patterns from large-messy-confusing data (Ho et al., 2021; Hua et al., 2020b; Pham et al., 2021). With considerable success in the computer vision domain, DL is now being leveraged in different domains, such as wireless communications, human–computer interaction, gaming, and finance. A few years ago, NVIDIA introduced DL super sampling (DLSS), a groundbreaking technology that exploits the power of DL and other AI algorithms to boost the frame rate while maintaining beautiful and sharp in-game images, thus being the potential to improve the visual experience in the metaverse. AI was also leveraged to improve game balance in several online multiplayer games by training supervised learning models iteratively until satisfying designers and play-testers. In some specific domains, the performance of AI-aided systems can be improved by various metaheuristic algorithms, such as prairie dog optimization algorithm (Ezugwu et al., 2022), dwarf mongoose optimization algorithm (Agushaka et al., 2022), and Ebola optimization search algorithm (Oyelade et al., 2022). To dive into a new era of 3D design simulation and collaboration for creating an impressive virtual-reality world, in the metaverse, as rich as the real world, NVIDIA introduced Omniverse,[1] an open-and-extensible platform, owning many valuable features, including physically-accurate simulation, multi-user design collaboration, photorealistic and real-time rendering, and AI-accelerated workflows.

---

[1] https://www.nvidia.com/en-us/omniverse/

1.1. Role of AI in the metaverse

By merging AI with other technologies, such as AR/VR, blockchain, and networking, the metaverse can create secure, scalable, and real-istic virtual worlds on a reliable and always-on platform. According to the seven-layer metaverse platform, it is undoubted to realize the important role of AI to guarantee the reliability of infrastructure and improve its performance so far. In the 5G and future 6G systems, many advanced ML algorithms with supervised learning and reinforcement learning have been adopted for different challenging tasks, such as efficient spectrum monitoring, automatic resource allocation, channel estimation, traffic off-loading, attack prevention, and network fault detection. With sensor-based wearable devices and other human–machine interaction gadgets, simple human movements and complex actions can be analyzed and recognized based on learning ML and DL models. Therefore, users' movements in the real world are projected into the virtual worlds, allowing users to fully control their avatars to interact with other objects in the metaverse comfortably. Moreover, these avatars can engage with many modalities adopted in the real world, such as facial expressions, emotions, body movement, and physical interactions, besides speech recognition and sentiment analysis, which are powered by AI in terms of accuracy and processing speed.

Although XR/VR someway represents the facade of a metaverse with immersive devices like head-mounted displays, AI is a pivotal technology working behind the scenes to build a creative and beautiful world, thus bringing a seamless virtual-reality experience to users. AI can facilitate the content creation process, for example, some AI modules like GANverse3D introduced by NVIDIA enable developers and creators to take photos of objects and then make virtual replicas. Several DL-based methods have been proposed for rendering 3D objects (including human body parts), which can achieve very impressive accuracy while presenting real-time processing accelerated by both software (e.g., PyTorch3D library from Facebook AI and TensorRT from NVIDIA) and hardware (e.g., GPUs) Meta just very recently introduced the AI research supercluster (RSC), believed as among the world-class fastest AI supercomputer that will speed up AI research and be served for building the metaverse (Moore, 2022). Furthermore, RSC can help AI researchers and scientists develop better DL models from massive data, including text, speech, image, video, for various services/applications. Accordingly, any achievements and outcomes derived from RSC will be used as fabrics to build the metaverse platform, in which AI-driven products will be of considerable importance.

1.2. Our contributions

The metaverse is built by merging advanced technologies to bring a completely 3D immersive experience to users, where they can truly interact and collaborate with others in the virtual worlds. Among such technologies like blockchain, XR/VR, and 5G, AI has a silent but important role in the foundation and development of the metaverse. However, understanding how AI can affect and contribute to the metaverse in the technical and application aspects is dubious, especially in the context of which it is neither mentioned in a fancy way like XR/VR nor discussed gloriously on social media like blockchain. No existing work is done to provide a comprehensive review of the role and use of AI in the metaverse. Consequently, we convey a comprehensive survey of the existing AI-based works from the technical and application perspectives; and further discuss their potentials for the metaverse. In a nutshell, the main contributions of our work are summarized as follows.

- We survey the state-of-the-art AI-powered approaches relying on various technical aspects, which show great potential for the foundation and development of a metaverse platform.
- A wide investigation of the existing AI-aided methods is provided for different applications that receive more interest to be deployed in the metaverse.
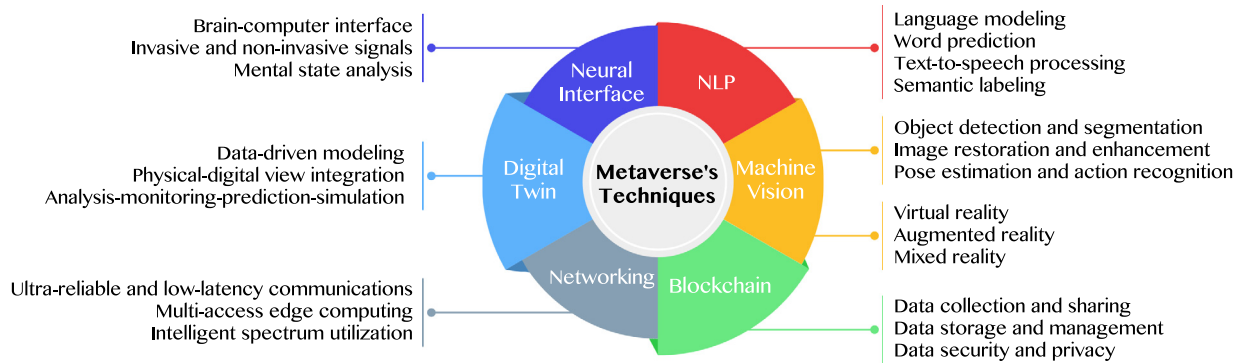
Fig. 4. Primary technical aspects in the metaverse, in which AI with ML algorithms and DL architectures is advancing the user experience in the virtual world.

- We introduce several interesting metaverse projects which have applied AI to enhance the immersive experience and develop user-oriented services. Moreover, some future research directions in AI for the metaverse are discussed.

The references reviewed and discussed in this work are found and obtained from high-reputed publishers, such as IEEE, Elsevier, Springer Nature, Wiley, Taylor & Francis, and well-known archival websites like arXiv. Furthermore, the following queries are used to find the related references, including "artificial intelligence", "machine learning", "deep learning", "natural language processing", "blockchain", "computer vision", "neural interface", "blockchain", "immersive technology", "virtual reality", "DL for healthcare and medical", "ML for manufacturing", "AI for gaming", and "metaverse projects". Our review protocol closely complies with the following process: criteria selection, quality assessment, data extraction, study synthesis, limitation disclosure, and reporting.

1.3. Paper organization

The remaining of our work is organized as follows. At first, the adoptions of AI in existing works that are the potentials for the foundation and development of the metaverse in various technical aspects are investigated in Section 2. Secondly, Section 3 explores several AI-powered applications to be promisingly deployed in the virtual worlds of the metaverse. Accordingly, some AI-aided metaverse projects are briefly carried out in Section 4. Finally, we conclude the paper with some future research directions for the development of the metaverse in Section 5.

2. AI for the metaverse: Technical aspect

This section investigates the state-of-the-art methods which have exploited AI, including traditional ML algorithms and DL architectures, to drive different metaverse enabling technologies: natural language processing, machine vision, blockchain, networking, digital twins, and neural interface as illustrated in Fig. 4. Although these works are not contributing to the metaverse directly, they show many potentials for the foundation and the development of virtual worlds in the metaverse via modernizing built-in services and applications. Accordingly, the immersive experience of users in the metaverse is enhanced significantly with nearly no boundary between the virtual world and the real world.

2.1. Natural language processing

Natural language processing (NPL), also known as computational linguistics, encompasses a variety of computational models and learning processes to solve practical problems of automatically analyzing and understanding human languages, including speech and text.

Besides, the field of NLS considers many topics, such as speech-to-text, text-to-speech, conversation design, voice branding, and multi-language and multi-cultural in voice. Furthermore, NLP plays a vital role in the metaverse regarding intelligent virtual assistants (a.k.a., chatbot). Particularly, NLP is principally responsible for enabling chatbots to understand complicated human conversation in the context of varying dialects and undertones. Empowered by AI, chatbots can answer nuanced questions and learn from interaction to improve the quality of responses. The AI chatbots are developed to assist users in some virtual environments like the metaverse.

As one of the most important tasks in NLP, language modeling predicts words or simple linguistic units by capturing syntactic and semantic relations of preceding words and units, which is useful for machine translation and text recommendation. In Daniluk et al. (2017), many neural networks with key–value attention mechanisms were built and evaluated on the Wikipedia corpus dataset to conclude that RNNs and LSTM networks with the attention mechanisms can outperforms large-scale networks while reducing memory in use. In Benes et al. (2017), a memory network with residual connection was designed to improve the performance of language modeling in terms of test perplexity if compared with regular LSTM having an equivalent size. Some recent CNNs have been leveraged to address the long-term dependencies in long sentences and short paragraphs, especially being efficient to specific and complicated word patterns (Pham et al., 2016). Some deep networks were designed with advanced modules and connection structures to enhance language modeling efficiency, such as gated connection and bi-directional structure (Liu and Yin, 2020). Besides word-aware language models, many character-aware models have been introduced with AI algorithms to deal with various diversified languages in the world. For an efficient analysis of the representation of words from characters, both the CNN and LSTM architectures were simultaneously applied in Athiwaratkun and Stokes (2017). Some models were evaluated on many datasets of English, German, Spanish, French, and Arabic, in which they showed the effectiveness in identifying prefixes and suffixes, recognizing hyphenated words, and detecting misspelled words (Ma et al., 2020; Sharma et al., 2020; Khysru et al., 2021). Generally speaking, character-aware and word-aware modeling techniques allow natural language understanding systems to extract syntactic and semantic information for some common tasks in the metaverse, such as part-of-speech tagging, named-entity recognition, and semantic role labeling.

DL has been further exploited to overcome the learning limitation of conventional ML algorithms and effectively deal with many challenging tasks in NLP. Some CNNs with simple and advanced architectures were leveraged to cope with multiple sentence-based tasks, such as sentiment prediction and question type classification (Jin et al., 2020b). Moreover, sentiment analysis and recognition tasks may require feature extraction of aspects and sentiment polarities, which are potential to improve the reliability and flexibility of virtual assistant units in the metaverse (Wang et al., 2019b). Natural language generation is an
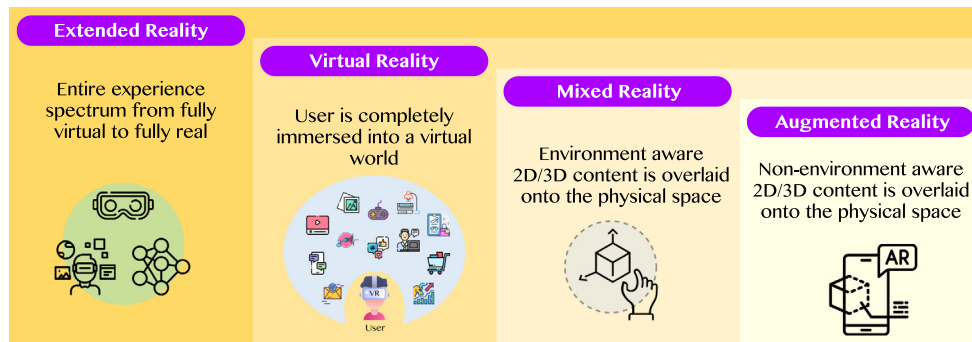
Fig. 5. The difference between AR, MR, and VR under the umbrella of XR.

advanced functionality of chatbot to generate reasonable task-specific conversation-oriented text. Some single RNN/LSTM and mixture LSTM-CNN models were proposed to generate short text in image captioning and long text in virtual question answer (Liu et al., 2019b). Besides supervised learning, unsupervised and reinforcement learning with deep models have been adopted for some specific NLP tasks, such as text parsing, semantic labeling, context retrieval, language interpretation, and dialogue generation (Young et al., 2018). In the metaverse, NLP techniques should be combined to fully provide text-based and speech-based interactive experiences between human users and virtual assistant.

2.2. Machine vision

Machine vision, including computer vision and XR in cooperation, is one of the central technologies to obtain the foundation of the metaverse. The raw data perceived from visual environments (via optical display and video player) is captured and processed to infer high-level information, which is then shown to users over head-mounted devices and others, such as smart glasses and smartphones. Indeed, computer vision allows XR devices to analyze and understand user activities based on visual-based meaningful information. Represented as avatars in the virtual worlds, the users can freely move in 3D maps and interact with virtual objects in the metaverse.

2.2.1. Extended reality

XR is defined as an umbrella term that encapsulates VR, AR, mixed reality (MR), and everything in their gaps as shown in Fig. 5. Although some revolutionary experiences are offered for VR and AR, the same original technologies are fueling the innovation and development of MR. While AR provides the experiences of graphics, video streams, and holograms in the physical world and VR offers viewing experiences in a fully immersive digital world, MR can deliver a transition experience between AR and VR. Along with these reality technologies, human users can experience the metaverse and enjoy diversified services in both the physical and digital worlds. While XR and AI are distinct sectors, they can be combined to reach a fully immersive in the metaverse.

While conventional two-dimensional (2D) videos are limited by the small field of view (FoV), the 360-degree videos providing unlimited view-point with all directions are suitable for VR performance. Many commercial VR headsets are designed to satisfy the high-class using requirements, such as performance and comfortableness, which encompass multiple tasks driven by AI automatically. With the VR headset, a human user can experience various services and applications in the metaverse, and further create hyperreal media contents in the virtual world. Some AI algorithms have been applied in VR devices to improve the human–machine interaction experience based on visual-based information. The work (Hu et al., 2021), for the prediction of user's eye fixations in some gaze-based applications, such as content design and rendering, a DL framework with multiple CNNs was built to deal with various kinds of input data, e.g., VR image, gaze data, and

head data. This model effectively exploited the correlations between eye fixations and other factors likes VR content and headset motion. In Pham (2018), neural networks were adopted for human identification and authentication by analyzing periodic behaviors between users and VR gears (e.g., controllers and head-mounted display). This work's effectiveness in delivering useful information and treatment recommendation was verified in collaboration and gaming scenarios and has shown some applicable potentials in other scenarios, such as working and shopping. To enhance users' quality of experience (QoE) in the virtual world, an innovative human–machine interface approach was proposed by incorporating triboelectric sensory gloves and display components in VR devices to recognize multi-dimensional motion of gestures (Zhu et al., 2020b). As a result, virtual objects which are recognized by leveraging ML/DL algorithms can be manipulated in the real-time VR/AR space. In Kataoka et al. (2019), not only AR headsets but also other devices (e.g., triboelectric gloves, hand-held touchscreen devices, and tabletops) are taken into consideration regarding specific applications, services, and infrastructures to access the contents in the metaverse and interact with virtual objects in the digital world.

As the satisfaction of high-resolution video viewing experiences with VR devices, it is necessary to develop an effective video quality assessment method, in which DL has represented as a powerful tool to obtain quantitative and qualitative benchmark objectives. In Wu et al. (2019), a high-performance method was developed for VR quality assessment by building a 3D CNN architecture, in which the video prediction results are validated via different widely-used image quality assessment metrics without video reference. Compared with the baseline, which performed handcrafted feature extraction and ML algorithms, the 3D CNN-based approach showed the superiority in term of VR video quality assessment and benchmark (Yang et al., 2018). To handle the limited data transmission bandwidth in VR systems, the work (Jin et al., 2021) extended the regular image quality assessment for both 2D and 3D foveated-compressed videos. The advantages and limitations of current video quality assessment methods were analyzed and exposed in Han et al. (2020), which can be useful to design an effective video transmission mechanism for various AR systems and diversified video contents. On the way to become the next mainstream for consumers and business, MR is defined as a blend of physical and digital worlds, which establishes the natural and intuitive interactions between 3D human, computer, and surrounding environment (Barba et al., 2012). This new reality is modernized by the recent revolutions in computer vision, graphic processing, display, remote sensing, and AI technologies (Kersten-Oertel et al., 2012). Compared with VR and AR, MR has more potentials for the metaverse thanks to its hybrid physical-virtual experiences via two main types of devices[2]: holographic device with see-through display allows users to manipulate physical objects while wearing it and immersive device allows users to interact with virtual objects in the virtual world. In the future, new devices to

---

[2] https://docs.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality

enhance the visual-interactive experiences of users in the metaverse should minimize the different gaps in terms of specification and utility between holographic devices and immersive devices.

2.2.2. Computer vision

In the last decades, computer vision has been empowered by AI, especially DL with a variety of network architectures to improve the overall accuracy of visual systems with efficient cost thanks to high-performance graphic processing units. Some fundamental computer vision technologies are potential to enhance the experience of human users in the metaverse, thus enabling users in the physical world to interact with the virtual environment in the digital world smoothly.

Semantic segmentation and object detection are two fundamental tasks in the computer vision domain, where semantic segmentation categorizes each pixel in an image to be one of pre-defined semantic classes and object detection aims to localize all possible objects in an input image by drawing corresponding bounding boxes with object information in tag (Liu et al., 2021a; Huynh-The et al., 2017). Early segmentation works mostly adopted local feature extraction and tracking in cooperation with classification algorithms, which were limited by unacceptable segmentation performance when dealing with large-messy datasets. Recently, numerous DL-based approaches have shown considerable improvement in terms of performance by exploiting different deep architectures if compared with traditional methods (Minaee et al., 2021). The powerful capability of CNNs in extracting deep visual features at multi-scale image resolutions has been exploited to design advanced segmentation models in Long et al. (2015), Shelhamer et al. (2017), Chen et al. (2018b), Wei et al. (2017) and Zhou et al. (2020). For example, DeepLab (Chen et al., 2018b) leveraged atrous convolution to improve feature learning efficiency by enlarging the receptive field of filters while keeping a small number of parameters or a low computational cost. Because of learning classification models at pixel level, image segmentation usually consumes considerable computation and large memory (Hua et al., 2020a, 2018). To overcome this challenge, some efforts have been presented with some skills of network designs and learning techniques, such as transfer learning (Liu et al., 2019a). The virtual environment in the metaverse is usually built with diversified visual units (e.g., single objects and multi-object modules); therefore, AI-based object detection must deal with a huge number of complicated classes, including real and virtual objects. Numerous recent object detection works have exploited CNN architectures to achieve impressive performance in terms of accuracy and processing speed (Cheng et al., 2018; Zhu et al., 2019; Liu et al., 2020a). DL-based semi-supervised and unsupervised learning models have been recommended to tackle unseen classes in the training dataset (Tang et al., 2018). Some natural problems of object detection in the 3D environment, such as occlusion, illumination variation, and view-point change, have been taken into consideration by incorporating advanced image processing and depth sensing algorithms (Chen et al., 2018a; Feng et al., 2021). In this context, depth estimation can improve the accuracy of object positioning in the 3D virtual world, but more geometric sensors are required to estimate depth information (Lai et al., 2019) (see Fig. 6).

In the virtual world, some image quality reduction problems, such as noise, blurring, and low resolution, should be addressed to enrich the visual perception of users. From the perspective of image processing and computer vision, these problems are studied over two tasks: image restoration and image enhancement. In Yeh et al. (2019), a decomposition-guided multi-scale CNN-based method was proposed to remove single image haze, in which deep residual structure and U-Net learning frame are combined to improve decomposed image components (so-called feature maps) while avoiding color distortion (Ronneberger et al., 2015). Some other advanced image restoration works exploited CNN architectures to reduce image compression artifacts, restore clean images from downscaled and blurred images, and reconstruct missing details (Wang et al., 2019a; Jin et al., 2020a; Zhang

et al., 2021). It is noted that the differences in terms of image quality and video specification between the clean virtual contents and real displayed images/videos can appear in VR devices (Wang, 2019; Saalfeld et al., 2021). These gaps can be filled effectively with AI-empowered image restoration methods, such as blur estimation, hazy removal, color correction, and texture reconstruction, but the computational complexity should satisfy the real-time video processing speed (usually measured by frames per second—FPS metric) to guarantee high-class user experience in the metaverse (Lahiri et al., 2020). Image enhancement has been widely considered for XR with some common tasks, such as contrast increment and super-resolution construction. In the past, many traditional image enhancement methods have been studied by applying image processing techniques, for example, histogram analysis and image decomposition (Huynh-The et al., 2014; Zhang et al., 2019). Recently, numerous impressive works of image enhancement have achieved considerable performance improvement by exploiting ML algorithms (Syrris et al., 2015) and DL with CNN architectures (Wang and Hu, 2020; Mei et al., 2020; Lee et al., 2020; Yang et al., 2020). For example, the work (Wang and Hu, 2020) introduced a convolutional down-sampling and up-sampling network to improve the overall contrast of images, in which the deep features of RGB (red, green, and blue) channels are combined via a feature-based fusion scheme to obtain cross-channel contrast balance. In Dong et al. (2016), a full CNN for image super-resolution was proposed with a lightweight structure, which can learn an end-to-end relation between input low-resolution images and output high-resolution images. Compared with some traditional sparse-coding-based methods in Yang et al. (2010), this approach has shown superiority in terms of image quality and processing speed. Super-resolution can become a cost-efficient solution that allows service providers to build high-resolution virtual worlds from low-resolution image/video sources.

In the metaverse, play users can control their avatars (or virtual characters) and interact with other users or non-player characters (NPCs), in which the posture and action of avatars should be estimated and recognized automatically with the support of motion-sensing interactive devices, such as controllers, gloves, and cameras (Dang et al., 2019). While human pose estimation aims to identify the body parts (or key body joints of a skeleton) and then track them in the real-time environment (Zheng et al., 2020), action recognition allows systems to understand single actions and complex interactive activities (e.g., human–machine interaction and human–human interaction) (Ramanathan et al., 2014). For instance, based on regular structural support vector regression (SVR) and latent structural SVR, two discriminative models in Chen et al. (2011) were proposed to deal with the problem of estimating human pose in cluttered environments. Interestingly, they are capable of capturing structural dependencies as the correlations between local features regarding pose representation. To improve the accuracy of body part localization and deal with varying views, the depth information acquired by depth cameras has been learned along with color information by advanced ML and DL models (Rogez et al., 2020; He et al., 2019; Zhao et al., 2021). In these works, some CNN architectures with high-class structural connections (e.g., dense layer connection, skip connection, and channel-attention connection) were designed to estimate skeletal joints precisely besides addressing some challenging problems in computer vision like object occlusion. In the line, human pose estimation has a close relation to action recognition, where the captured body information is useful to identify actions via pattern recognition models. Instead of detecting instant postures that can expose high confusion, many current works have tracked body motion in the temporal domain for long-term observation to improve the accuracy of action recognition. For example, in Huynh-The et al. (2016), Tu et al. (2019) and Huynh-The et al. (2018), some generative statistic models have been developed to analyze human pose transition by capturing the spatiotemporal geometric features between different body parts. Notably, the last decade has witnessed the revolution of visual-based action recognition with DL
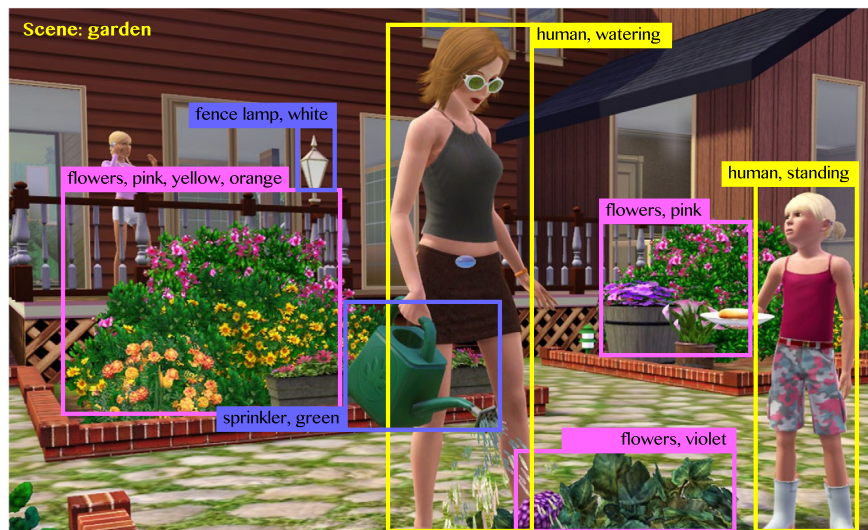
Fig. 6. Computer vision in the metaverse with scene understanding, object detection, and human action/activity recognition.

to significantly improve recognition accuracy and effectively deal with numerous realistic single actions and grouped activities (Wang et al., 2017). Some methods proposed innovative networks with advanced CNN architectures (Huynh-The et al., 2020a) and hybrid CNN-RNN architectures (Wang et al., 2017) to improve the learning efficiency of action discrimination models. Additionally, hand gesture recognition, gait identification, and eye tracking have been considered to improve interactive experiences in XR environments (Huynh-The et al., 2020c; Hu et al., 2020).

2.3. Blockchain

In general, blockchain is defined as a digital ledger that contains a list of recorded transactions and tracked assets interconnected in a business network by using cryptography techniques (Maddikunta et al., 2022). Blockchain can provide immediate, shared, and transparent information stored in an immutable and impenetrable ledger which can be accessed by only the network members with permission (Gadekallu et al., 2022). A typical blockchain network can track orders, payments, accounts, and other transactions. In the metaverse, a large amount of data (e.g., videos and other digital contents) is acquired by VR devices, transmitted over networks, and stored in data center without any security and privacy protection mechanisms, which can become the sensitive target of cyberattacks. In this context, blockchain with several unique features reveals a promising solution for security and privacy issues in the metaverse (Cannavò and Lamberti, 2021), especially when it is empowered by AI technologies. Besides, many creative activities and events offered by service providers to users will yield numerous in-metaverse objects/items (a.k.a., digital assets) which should be recorded and tracked via transparent transactions with smart contracts in blockchain.

In the last decade, numerous advanced data acquisition and storage methods have been proposed by combining blockchain and AI technologies in various application domains to improve data security and privacy, which have shown great potential to be deployed in the metaverse (Liu et al., 2020b). For instance, in Tanwar et al. (2019), various conventional ML algorithms (e.g., clustering, support vector machine (SVM), and bagging) and innovative DL architectures (e.g., CNN and LSTM) were investigated for data analytics to detect and classify cyberattacks in blockchain-based data acquisition and storage. Some other concerns were also considered in this work, such as incentive mechanisms to encourage users to contribute authenticated data, AI-based smart contract evaluation, and cost-efficient model learning in on-chain environment. For the Internet of Things (IoT)-aided smart cities, the

work (Kumar et al., 2021) introduced an effective privacy-preserving and secure framework by integrating blockchain with enhanced proof of work and ML with data transformation, which in turn robustly deals with various cyberattacks during data acquisition in smart city networks. In Khan et al. (2021), deep extreme learning machine was exploited in a resource-efficient blockchain-based IoT framework, which improved the system security and privacy based on data interpretation and abnormality prediction. This framework (see Fig. 7) has shown high performance of fraud detection and threat prediction, and can be extended for dealing with security and privacy problems in data storage and sharing instead of data collection. Recently, DL has replaced traditional ML in terms of cooperating with blockchain to solve some challenging security and privacy issues of big data, where five essential characteristics of big data (i.e., velocity, volume, value, variety, and veracity) are presented. For example, DeepChain Weng et al. (2021), a CNN-based blockchain framework, was developed to ensure the privacy and integrity of data contributed by participants in a network. Deep RL was exploited to achieve a secure mobile offloading in multi-access edge computing (MEC) based blockchain networks and to obtain a secure vehicular crowdsensing in the blockchain-based Internet of vehicles (IoV) systems (Nguyen et al., 2020; Wang et al., 2021b).

Federated learning (FL) has recently emerged as an effective solution to address the privacy problems of data sharing, in which multiple users train AI models with their own local data and collaboratively learn a global model at the server by a parameter aggregation mechanism (Pham et al., 2022; Ramu et al., 2022). In Lu et al. (2020), FL was applied to address privacy issues in data sharing among multiple untrusted parties in a blockchain network. This work integrates FL into a proof of training quality, a novel consensus mechanism, of permissioned blockchain to reduce computing and communication costs. To guarantee high privacy of massive data generated by heterogeneous IoT devices, FL was deployed in a blockchain-based resource trading system (Fan et al., 2021). A smart contract-based incentive algorithm was proposed to encourage edge nodes to contribute and evaluate FL tasks. For vehicular edge computing in intelligent transportation systems, FL was combined with blockchain to collaboratively detect malicious attacks (Liu et al., 2021b). While FL can offload the trained intrusion detection model to distributed edge devices to reduce computing resources of the central server, blockchain can ensure the security of the aggregation model in both the model storage and sharing processes. Besides data security and privacy, interoperability is another important concern in blockchain to collaborate with different parties using different data infrastructures. For example, a learning analytics framework (Ocheja et al., 2018) was studied to obtain solid interoperability between multiple blockchain participants which have to share
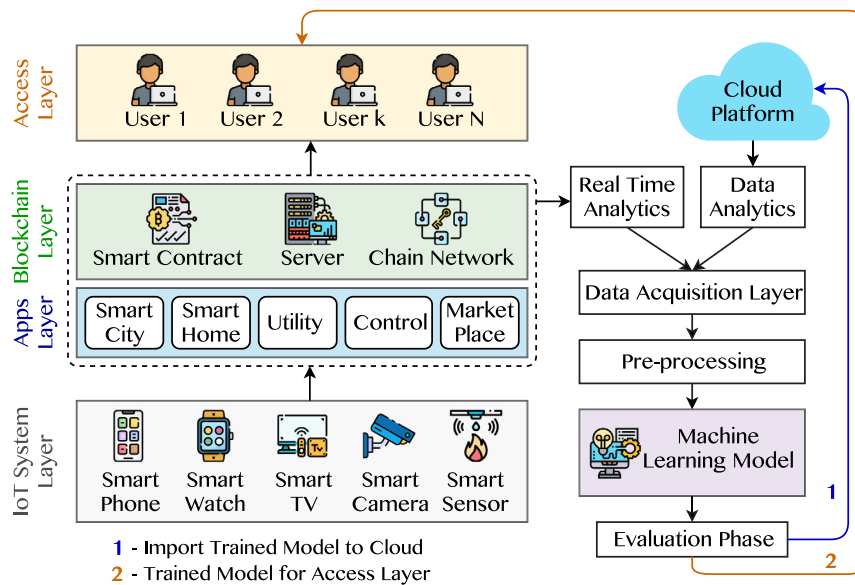
Fig. 7. A blockchain-based IoT framework with ML to enhance security and privacy.

a single ledger. Integrating blockchain into FL were recently found in computing resource allocation and management applications to address various problematic issues in centralized systems, such as external cyberattacks, server malfunctions, and untrustworthy server (Majeed et al., 2021; Li et al., 2021). In the metaverse, wherein multiple parties join and contribute digital content having different formats and structures, data security, privacy, and interoperability can be fully handled by collaboratively developing blockchain and AI.

2.4. Networking

The metaverse serves a massive number of users regarding pervasive network access over wireless networks. In the last decade, several innovative technologies have been introduced to improve the overall performance of wireless communication and networking systems, in which AI has been intensively used at multiple layers of a network architecture (Chen et al., 2019). Real-time multimedia services and applications in the metaverse usually demand a reliable connection with high throughput and low latency to guarantee a basic-level user experience at least. As the requirements of fifth-generation (5G) networks, the peak data rate should be around 10 Gbps (gigabits per second) and the end-to-end delay cannot exceed 10 ms (millisecond). In this context, ultra-reliable and low-latency communications (uRLLC) represent the foundation to enable the development of emerging mission-critical applications (She et al., 2021). Several optimization algorithms have been introduced to achieve uRLLC in 5G networks and beyond (e.g., sixth-generation 6G), but most of them require high computing resources. ML and DL have shown great potential to effectively handle existing challenging tasks, such as intelligent radio resource allocation (She et al., 2020), in 5G/6G networks while meeting a very low latency. RL was leveraged to address the resource slicing problem for enhanced mobile broadband (eMBB) and uRLLC, in which the complicated patterns of resource allocation and scheduling are formulated to collaboratively learn network states and channel conditions (Alsenwi et al., 2021). RL also showed the effectiveness in terms of joint subcarrier-power management and allocation, thus significantly reducing latency and improving reliability on the Internet of controllable things in Gu et al. (2021). Particularly, this work proposed a double Q-learning network to optimize the total spectrum efficiency via subcarrier assignment and power control policy, and accelerate learning convergence. As a vital role to enable uRLLC, efficient radio resource management was investigated with a distributed risk-aware ML approach to monitor

and manage the transmission of non-scheduled and scheduled uRLLC traffics (Azari et al., 2019).

Lately, DL has been exploited for many tasks in uRLLC, including spectrum management, channel prediction, traffic estimation, and mobility prediction. Two advanced CNN architectures, namely MC-Net (Huynh-The et al., 2020b) and SCGNet (Tunze et al., 2020), were designed in physical layer to automatically identify the modulation types of incoming signals, which in turn allows the receiver to demodulate accurately and enhance the spectrum utilization efficiency accordingly. To overcome the high computational cost of conventional channel state information (CSI) estimation approach, an online CSI prediction method was proposed a supervised learning framework by combining CNN and LSTM, in which two-stage training mechanism was deployed to improve the robustness and stableness of CSI estimation in practical 5G wireless systems (Luo et al., 2020). In Guo et al. (2019b), an end-to-end CNN architecture was designed with 3D convolution for intelligent cellular traffic forecasting, in which the deep model can learn the underlying correlations of traffic data in both the short-term and long-term spatial patterns. Besides achieving high accuracy of traffic prediction, the deep network showed the effectiveness with different real-world scenarios, such as traffic congestion data and crowd flow data. In conclusion, with ML and DL architectures, AI is a powerful tool to address many challenging problems of uRLLC in future wireless networks, which allows users to experience high-class integrated services in the metaverse with the guarantee of high throughput low latency (see Fig. 8).

2.5. Digital twins

As a digital representation of real-world entities, a DT can synchronize operational assets, processes, and systems with the real world along with some other regular actions, such as monitoring, visualizing, analyzing, and predicting (Tao et al., 2019). DTs are at the central of where the physical world and the virtual world interact via IoT connections; and therefore, any change in the real world will be rejected in the digital representation (Chen et al., 2021). With these distinctive properties, DT is found as one the fundamental building sectors of the metaverse and plays as the gateway for users to enter and enjoy services in the virtual world by creating exact replications of reality, including structure and functionality. For example, technicians can maneuver 3D representations of complex systems at multi-level sophistications (i.e., descriptive, informative, predictive, comprehensive, and
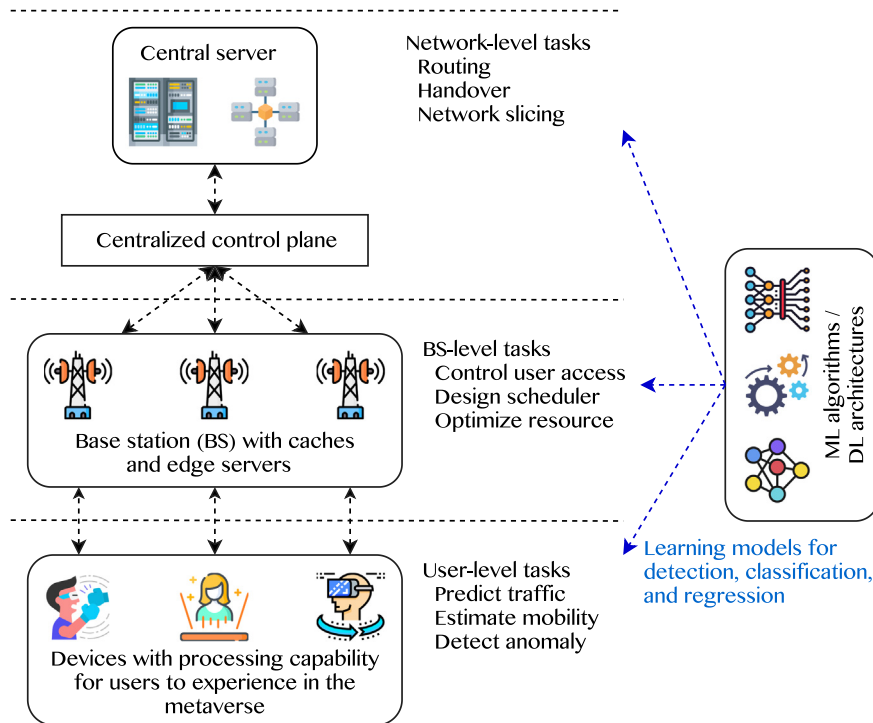
Fig. 8. A general architecture in 5G and beyond for metaverse services and applications, in which AI with ML algorithms and DL models contribute in multi-level tasks.

autonomous) for a wide spectrum of purposes, such as technical training and commercial customization. Accordingly, DTs allow application developers and service providers to reconstruct virtual replications of machines and processes, in which any kind of physical analysis can be done remotely with AI (Rathore et al., 2021).

For industry 4.0, a reliable DT framework (Darvishi et al., 2021) was proposed for sensor-fault detection, isolation, and accommodation, in which a multi-purpose ML method with multi-layer perception neural network was deployed to validate sensory data, estimate fault condition, and identify fault sensor. As a digital replication to operate human–robot welding actions, a DT system in Wang et al. (2021a) was developed along with VR and AI technologies to monitor and analyze welder behaviors. A generic ML framework with domain transformation, feature engineering, and classification was applied to recognize proper welding behaviors based on the data acquired from a bi-directional flow between robot and VR. In Elayan et al. (2021), a data-driven-based DT framework (see Fig. 9) was studied to improve health diagnosis performance and promote better health operation in intelligent healthcare systems. DTs contributed in different phases to create virtual replications of patent health profiles, carry out collaborative activities of health professionals, and formulate a universal treatment plan for patients in same cases. ML models were built to learn meaningful information from raw data collected by the Internet of Medical Things (IoMT) devices to early detect health abnormalities and precisely recognize health problems. In another works (Ghandar et al., 2021) that proposed a cyber–physical framework for smart urban agriculture services and application, DTs were designed to replicate a virtual representation of farming production, in which the sensory data acquired by practical sensors was processed by ML algorithms in a decision support system. To adapt to different types of product magnification, the DTs were built from small functioning modules to a whole process twin.

With a great capability of automatically learning features from high-dimensional unstructured data and effectively dealing with spatiotemporal learning models, DL has been recently applied in DT architectures for different services and applications. In Xu et al. (2022), a DT architecture was developed for the edge computing-aided Internet of

vehicles (IoV) to improve the utilization efficiency of the vehicle's computational resources. To overcome the overload problem of edge devices, a deep Q-network optimized the function approximation of DL and RL. For the performance investigation of uRLLC services and delay tolerant services in mobile edge computing systems (Dong et al., 2019), a DT framework was developed by replicating a virtual pattern of the real network environment. Remarkably, DL with feedforward neural network architecture was carried out to deal with varying network parameters of real-world phenomena. For industrial IoT, DTs were built to simulate and capture the operation state and real-time behavior of industrial devices, which were mapped into a digital world (Sun et al., 2021). To address the bias between the real entity and its digital replication, a trusted-based aggregation with FL was carried out with a deep RL model to generally improve performance while meeting resource constraints. With AI as a powerful analytics tool, DT can improve system performance, reduce process-related incidents, minimize maintenance costs, and optimize business and production. In addition, DT allows users to view the metaverse as an advancing replication of reality with full real-time synchronization from the physical world.

2.6. Neural interface

Technology is enriching the world around us by enhancing the human experience and filling the gap between reality and the virtual world in the metaverse. In this context, the most immersive popular interface to interact with virtual world is a VR headset with a controller. Many technology companies currently pay attention to neural interfaces, so-called brain–machine interfaces (BMIs) or brain–computer interfaces (BCI), that go beyond VR devices. The BMIs help to nearly clean the borderline between human and wearable devices. Many BMIs detect neural signals using external electrodes or optical sensors that adhere to the skull and other parts of the human body. According to these noninvasive devices, which only read and control the mind at a rudimentary level, BMIs can manipulate thoughts with transcranial electromagnetic pulses. Fig. 10 describes a common BMI cycle with primary components for processing neural signals and responding neural stimulations (Bernal et al., 2022). Besides data engineering techniques
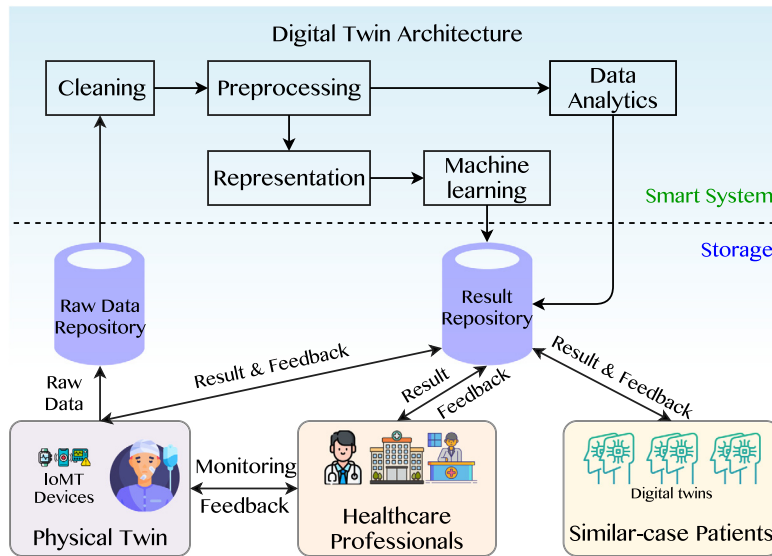
Fig. 9. A data-driven DT framework for intelligent healthcare systems using ML to process raw data of IoMT devices.
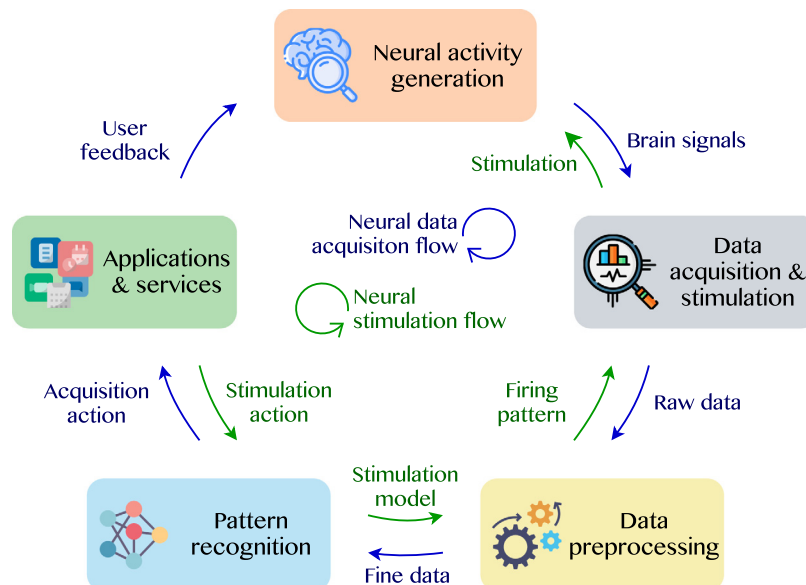

Fig. 10. A common BMI cycle with primary components for processing neural signals and responding neural stimulations.

in the preprocessing stage, AI/ML algorithms in the pattern recognition stage enable the analysis of complicated and sensitive neural signals accurately.

With electroencephalogram (EEG) signal as one of the most popular inputs of BCI systems, the work in He and Wu (2020) studied a brain signal classification by two learning approaches: one is offline unsupervised classification and another is simulated online supervised classification. Besides that, two approaches achieved lower computational cost and better performance in common tasks, e.g., motor imagery, mental analysis, and event-related potential, the offline unsupervised mechanism did not require signal labeling for new subjects in the learning phase. In building an accurate predictive model in BCIs to decipher brain activities into communication and control commands, the work (Abibullaev and Zollanvari, 2019) learned discriminative spatiotemporal features to seize the most relevant correlations between different neural activities from EEG signals. Based on the reconstructed signal waveforms containing dominant frequency characteristics as feature vectors with lower dimensionality, several ML algorithms, e.g., logistic regression (LR), Naive Bayes, and SVM, were

applied to investigate the performance of ERP. In Matran-Fernandez and Poli (2017), the feasibility of using the visual hemisphere to extract relevant information about the spatial location of targets in aerial images was investigated with feature selection and SVM classification, which were deployed in a rapid serial visual presentation (RSVP) procedure. Concretely, by learning ERP patterns from extracted discriminative features of EEG signals, the concerning target and its location in aerial images can be identified. As an effort to increase the correct classification rate of EEG signals in BCIs, an advanced ML framework was proposed by combining an improved common spatial pattern algorithm and a transfer learning mechanism (Lv et al., 2021). Besides achieving high accuracy in classifying left-hand and right-hand imaginary movements, the trained AI model can be used for other classification and recognition tasks in the same domain via the knowledge transferring technique.

Based on the superiority of capsule network (CapsNet) compared with traditional neural networks in terms of feature extraction and feature explanation, the work in Ma et al. (2021) applied CapsNet to improve the accuracy of ERP detection in BCI systems. With highly
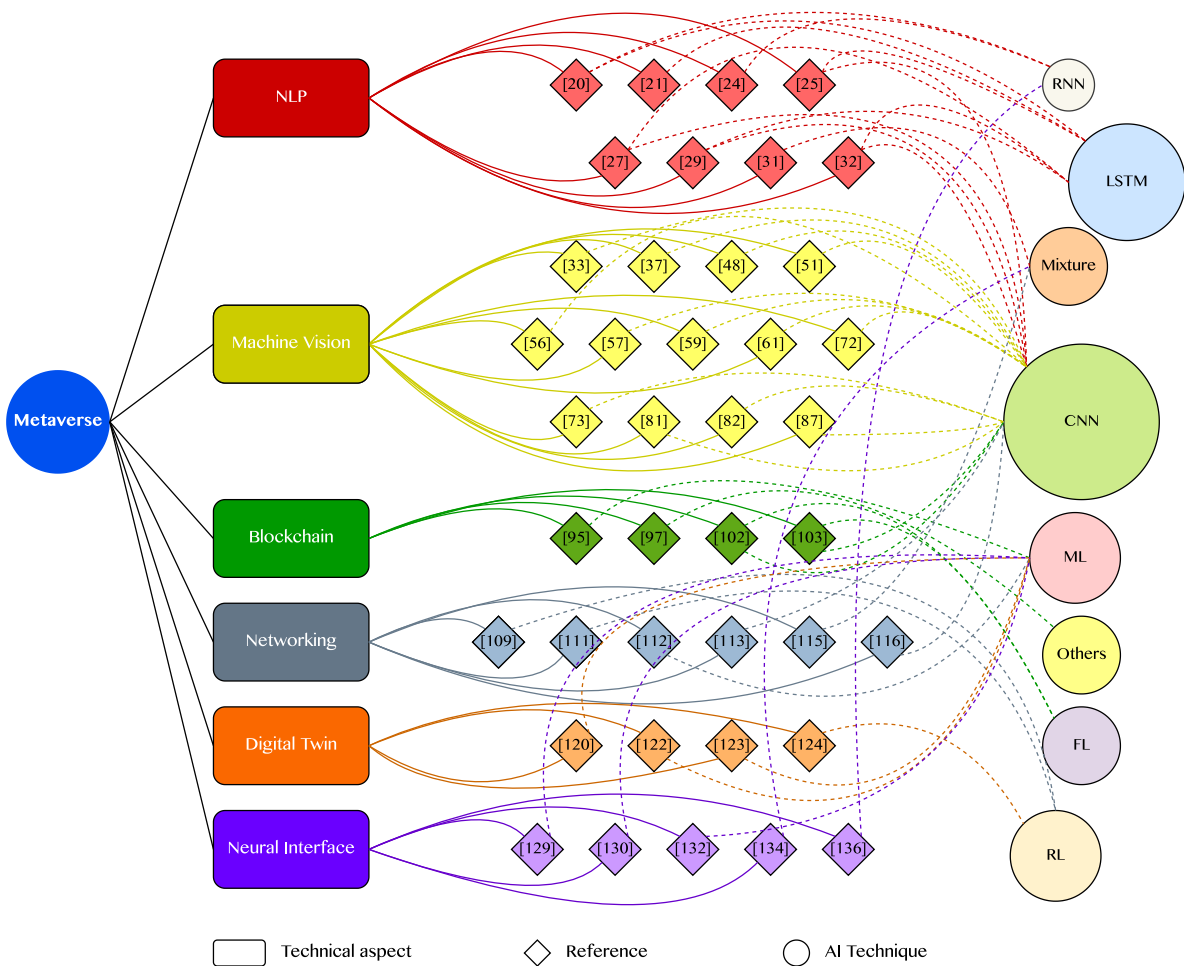
Fig. 11. A bibliometric analysis and visualization of potential AI-based technical works for the metaverse.

discriminative spatial features and key temporal correlations extracted from EEG signals by capsule layers, CapsNet not only outperformed some state-of-the-art learning models (e.g., linear discriminant analysis and CNN Sakhavi et al., 2018) but also obtained practicality with different common spellers in the cognitive neuroscience domain. In Jeong et al. (2020), a hybrid DL framework was proposed with multi-directional CNN and bidirectional LSTM in an accurate brain-controlled robotic arm system. This proposed learning approach effectively calculated underlying spatial signal correlations in time, boosting the decoding performance for 3D multi-directional arm-based object grasping tasks. Inspired by GoogleNet (Szegedy et al., 2015), the work (Santamaría-Vázquez et al., 2020) proposed EEG-Inception, a novel CNN for EEG-based classification tasks in BCI systems, which involved multiple inception modules in improving feature learning efficiency. Furthermore, an effective training strategy that incorporated cross-subject transfer learning and fine-tuning to reduce the calibration time of ERPs and demonstrate the feasibility of real-world assistive applications. In the future, brain–computer interfaces will truly promote the ultimate immersive interaction between reality and the virtual world in the metaverse via consumer-ready mind-control systems. The existing AI-enabled works on the subject of six concerning technical aspects, which are promisingly for the metaverse, are summarized in Table 1 and visualized by a bibliometric analysis graph in Fig. 11.

3. AI for the metaverse: Application aspect

This section conveys the existing AI-aided works in the perspectives of four key applications: healthcare, manufacturing, smart cities, and gaming (see Fig. 12); which are probably considered to delivery

specialized services in the metaverse. Besides, some other potential applications, including E-commerce, human resources, real estate, and decentralized finance, are discussed in brief.

3.1. Healthcare

The health industry has recently started exploiting some revolutionary techniques like VR and big data incorporated with AI in software and hardware to increase the proficiency of medical devices, reduce the cost of health services, improve healthcare operations, and expand the reach of medical care. From a 2D environment to a 3D virtual world, the metaverse allows users to learn, understand, and share patients' health conditions and medical reports in an immersive manner. By means of VR/XR systems, AI plays a vital role in many healthcare and medical sectors, for example, achieving better efficiency in providing diagnosis, delivering accurate and faster medical decisions, providing better real-time medical imaging and radiology, and supporting more convenient simulated environments to educate interns and medical students.

In many wearable devices for healthcare and wellness applications and services, AI has been applied to automatically recognize complex patterns of sensory data (Banos et al., 2015). In Huynh-The et al. (2021), a physical activity recognition method was introduced to support physicians and health-wellness experts to make decisions in daily living assistance and early health risk awareness by using the sensory data of multiple wearable devices. The method combined the globally handcrafted features and locally deep features (i.e., extracted by a deep CNN) over an intermediate fusion mechanism to improve the activity recognition rate. In Hur et al. (2018), a novel encoding

Table 1
Summary of AI-aided technologies being potential for developing the metaverse.

| Tech. Aspect	Ref	Task	AI Technique
NLP	Daniluk et al. (2017)	Word and linguistic prediction for language modeling.	RNNs and LSTM networks with the attention mechanisms.
	Benes et al. (2017)		Advanced memory network with residual connection.
	Liu and Yin (2020)		RNNs with gated connection and bi-directional structure.
	Athiwaratkun and Stokes (2017)	Understanding words from characters	General deep networks with CNN and LSTM architectures.
	Sharma et al. (2020)	Identifying prefixes/suffixes and misspelled words	DL framework with CNN, Bi-LSTM, and conditional random field.
	Jin et al. (2020b)	Sentiment prediction and question type classification.	Various CNNs and LSTM networks with simple structures and advanced-designed architectures.
	Liu et al. (2019b)	Generate short text in image captioning and long text in virtual question answer.	DL framework with single RNN/LSTM and mixture LSTM-CNN models.
	Young et al. (2018)	Semantic labeling, context retrieval, and language interpretation.	Unsupervised and reinforcement learning with common RNN/LSTM and CNN models.
	AI-powered NLP revolutionize intelligent virtual assistants to understand and respond complicated human conversations, thus significantly improving text-based and speech-based interactive experiences between users and virtual assistants in the metaverse.		
Vision Machine	Hu et al. (2021)	Forecasting eye fixations in VR-based environment.	DL framework with multiple CNNs for different tasks.
	Wu et al. (2019)	VR quality assessment for 2D and 3D videos	CNN-based DL framework with 3D convolutional layers.
	Chen et al. (2018b)	Semantic segmentation and object detection.	CNNs with atrous convolution.
	Hua et al. (2020a)		CNNs with channel-wisely and spatially attentional schemes.
	Liu et al. (2020a)		CNNs with pixel-wise local and global attention mechanims.
	Tang et al. (2018)		CNNs with similarity- and semantic-based knowledge transfer.
	Feng et al. (2021)		CNNs with 3D object-object relation graphs.
	Yeh et al. (2019)	Image/video quality enhancement (e.g., hazy removal, color correction, texture reconstruction, and super-resolution)	Decomposition-guided multi-scale CNN architectures with deep residual structure and U-Net learning frame.
	Wang and Hu (2020)		Combination of deep features via a feature-based fusion scheme.
	Mei et al. (2020)		3D CNN with simultaneous and separated feature learning.
	Chen et al. (2011)	Human pose estimation and action/activity recognition.	Discriminative model with latent structural SVRs.
	Rogez et al. (2020)		CNNs with dense layer and channel-attention connections.
	Huynh-The et al. (2018)		Generative models with Dirichlet and Pachinko allocations.
	Computer vision and XR technologies can help to build more beautiful 3D virtual worlds and enhance high-quality high-resolution video viewing experiences over different immersive devices.		
Blockchain	Tanwar et al. (2019)	Identifying cyberattacks in blockchain-aided networks.	Conventional ML algorithms and regular DL architectures.
	Khan et al. (2021)	Resource management in blockchain-based IoT.	Deep extreme learning machine.
	Fan et al. (2021)	Privacy preservation for heterogeneous IoT devices.	FL framework with deep model using average aggregation.
	Liu et al. (2021b)	Malicious attacks detection in transportation systems.	FL framework with CNN model averaging and training.
	Blockchains with the transparency and immutability features is realized as a promising solution to many security and privacy problems, in which AI helps to detect cyberattacks for protecting digital assets in the metaverse.		
Networking	She et al. (2020)	Resource slicing problem for eMBB and uRLLC.	RL with a policy gradient based actor-critic learning mechanism.
	Gu et al. (2021)	Subcarrier-power management and allocation.	RL with double Q-learning network.
	Azari et al. (2019)	Management of the transmission of non-scheduled and schedule uRLLC traffics.	Supervised learning framework with conventional ML algorithms.
	Huynh-The et al. (2020b)	Modulation classification in wireless systems.	Advanced CNN architecture with sophisticated-designed modules.
	Luo et al. (2020)	Prediction of CSI in 5G wireless systems.	Supervised learning framework by combining CNN and LSTM with two-stage training mechanism
	Guo et al. (2019b)	Forecasting intelligent cellular traffic.	End-to-end CNN architecture was designed with 3D convolution.
	AI-aided wireless communication and networking enables users to experience numerous high-class integrated services in the metaverse, wherein high throughput and low latency are guaranteed for heavy data transfer.		
Digital Twin	Darvishi et al. (2021)	Sensor-fault detection, isolation, and accommodation	Multi-purpose ML method with multi-layer perception networks.
	Elayan et al. (2021)	Diagnosis and detection of heart disease.	Data-driven-based analysis framework with ML algorithms.
	Ghandar et al. (2021)	Prediction of system states with DT of urban farming.	Several common ML algorithms for regression and classification.
	Xu et al. (2022)	Improvement of resource utilization in edge computing-aided IoVs network.	Deep Q-network to optimize the function approximations of DL and RL.
	DT allows users to perceive the metaverse as an advanced replication of reality with full operational features and functionalities synchronized from the physical world in real time.		
Neural Interface	Abibullaev and Zollanvari (2019)	ERP classification in BCI systems.	Spatiotemporal feature extraction and ML algorithms.
	Matran-Fernandez and Poli (2017)	Spatial object localization in aerial images.	Feature selection with SVM classification.
	Ma et al. (2021)	ERP detection in BCI systems.	Capsule network with primary capsule components.
	Jeong et al. (2020)	Brain-controlled robotic arm system.	DL with multi-directional CNN and bidirectional LSTM.
	Santamaría-Vázquez et al. (2020)	EEG-based classification tasks in BCI systems.	DL with cross-subject transfer learning and fine-tuning.
	AI-aided BCI systems truly stimulate the ground-breaking immersive interaction between reality and the virtual world in the metaverse via consumer-ready mind-control systems.		
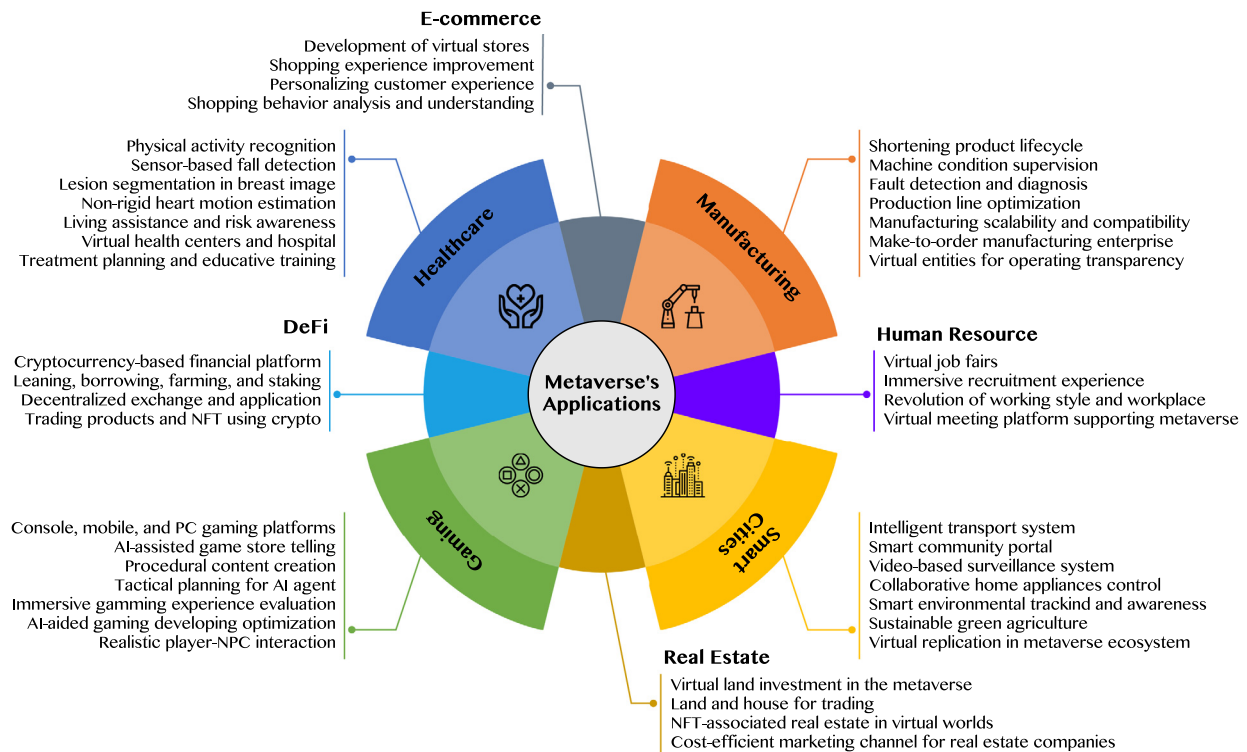
Fig. 12. AI for the metaverse in the application aspects with healthcare, manufacturing, smart cities, and gaming besides other promising domains, such as E-commerce, human resources, real estate, and DeFi.

algorithm, namely Iss2Image, was introduced to convert inertial sensory signals (e.g., accelerometer, gyroscope, and magnetometer) into a color image for CNN-based human activity classification. Furthermore, a lightweight CNN with few layers in a cascade connection was designed to learn the physical activity patterns from encoded activity images. The work (Qian et al., 2020) proposed a system of fall detection using wearable devices for IoT-based healthcare services, in which a hierarchical DL framework with CNN architectures was developed for collaboratively processing sensory data at local devices and a cloud server. As capable of working with multiple wearable devices (e.g., smartphones, smartwatches, and smart insoles), the system yielded a high correct detection rate with high data privacy. Besides CNN, RNN and LSTM networks were exploited to process wearable sensory data in some early healthy risk attentions, such as fall detection and heart failure (Li et al., 2020).

With the great success of DL, especially CNN architectures, in the image processing and computer vision domains, few recent years have witnessed a vast emergence of DL to address various challenging tasks of medical image analysis because of the requirement of much more specialized knowledge from technicians and medical experts if compared with natural image analysis (Hua et al., 2019, 2021). For lesion segmentation in breast ultrasound (BUS) images, the work (Ning et al., 2022) studied an advanced network, namely saliency-guided morphology-aware U-Net (SMU-Net), by involving an additional middle feature learning stream and an auxiliary network. The coarse-to-fine representative features from the auxiliary network were fused with other features (e.g., background-assisted, shape-aware, edge-aware, and position-aware) to effectively discriminate morphological texture in BUS images. In Qi et al. (2021), a cost-efficient unsupervised DL approach was introduced to accelerate the processing speed of non-rigid motion estimation of the heart in free-breathing 3D coronary magnetic resonance angiography images. Relying on a deep encoder–decoder architecture, the network can learn image similarity and motion smoothness without ground truth information in a patch-wise manner to save computing resources significantly instead of a regular

volume-wise manner. To overcome the obstacle of increasing network size and computation of 3D CNNs in mining complicated patterns in 3D images, 2D neuroevolutionary networks were investigated for 3D medical image segmentation, in which an optimal evolutionary 3D CNN was renovated to reduce computational cost without sacrificing accuracy (Ning et al., 2022; Hassanzadeh et al., 2021). With AI in use as the core technology for data analytics, several healthcare applications and medical diagnosis services (e.g., motor rehabilitation and magnetic resonance imaging neurofeedback) can be developed in the VR environment for multipurpose, such as collaborative treatment planning and educative training (Torner et al., 2019). Indeed, several healthcare applications and medical services can be provided in the metaverse. For example, medical students can improve surgical skills by doing interactive practice lessons built for medical education in the virtual world or patients can find some healthcare services via virtual assistants at virtual health centers and hospitals.

3.2. Manufacturing

As the current wave of the industrial revolution, digital transformation in manufacturing has been happening with the digital connection between machines and systems to better analyze and understand physical entities. Different from digital transformation to enhance the physical world via digital operations, the metaverse creates a virtual world that is translated onto the physical world based on the foundation of real interaction and persistence. By collaboratively adopting cutting-edge technologies, such as AI and DT, the metaverse for manufacturing can significantly modernize digital operations in the current digital revolution. Currently, AI with ML algorithms and DL architectures have considerably contributed to the manufacturing domain via numerous industrial applications (Boopalan et al., 2022).

In manufacturing, shortening product lifecycles and increasing the number of product variants are the main reason for high expense for frequent production system reconfigurations and upgrades, especially with ML-based systems which have spent more time and computing

resources on new data collection, preprocessing, and model learning. To overcome the above challenges, a symbiotic human-ML framework was leveraged with a reinforcement learning strategy by combining the learning capacity of Q-learning models and the domain knowledge of experts (Doltsinis et al., 2018). This framework also considered human exploration to reduce noise in data and improve the quality of automatic decision-making systems. As of great importance in the modern manufacturing systems, quality inspection has been recently attracting more attention with intelligent data-driven condition supervision approaches; however, they had to face some difficulties from different operating conditions with diversified tasks and applications (Azamfar et al., 2020). For reliable fault detection and diagnosis in manufacturing, numerous methods have exploited DL with RNN and CNN architectures to achieve high accuracy while keeping real-time monitoring. For instance, an RNN (Lee et al., 2021) was developed with an encoder–decoder structure coupled with an attention mechanism to predict and diagnose inter-turn short-circuit faults in permanent magnet synchronous systems. In Xue et al. (2020), a data-driven LSTM-based fault diagnosis approach was introduced to early detect multiple open-circuit faults in wind turbine systems. In Guo et al. (2019a), a DL-based intelligent fault diagnosis method was introduced to address two challenging problems, i.e., the lack of labeled data for the learning model and the data distribution discrepancy between training and testing sets, by incorporating CNN architecture and transfer learning mechanism.

Design and implementation of an optimal serial production line are crucial to increase the productivity of the whole manufacturing process. Many recent works have applied AI to optimize some specific sectors in a production system and improve the performance of the production line accordingly while meeting scalability and compatibility. For example, a prediction model (Alkhalefah et al., 2021) was developed to estimate the optimal buffer size in production lines by combining a regular artificial neural network (ANN) and a genetic algorithm. The prediction model was further integrated with an optimization mechanism to evaluate and forecast the optimal buffer size in need to maximize productivity. In Huang et al. (2019), an efficient production progress prediction method was formulated with the combination of DL and IoT to optimize the dynamic production and on-time order delivery activities in make-to-order manufacturing enterprises. In the proposed method, a two-stage transfer learning mechanism was executed with historical data and real-time state data using a deep belief network to solve the nonlinearity of production progress. Nowadays many manufacturing plants have developed industrial collaborative robots to undertake different advanced tasks which require much more cognitive skills, intelligence, and domain knowledge of humans to immediately respond to unexpected actions or events with high precision and confidence (Lins and Givigi, 2021). Therefore, it demands a cooperative AI model to learn complicated patterns from multimodal data for different correlated tasks of the manufacturing process and production line, in which the AI model should be equipped with the capability of explanation and reasoning. Through virtual entities in the metaverse, industrial manufacturing efficiency is generally improved with AI to speed up production process design, motivate collaborative product development, reduce operation risk to quality control, and obtain high transparency for producers and customers.

3.3. Smart cities

Smart cities acquire the meaningful information about the needs of citizens through the IoT, video cameras, social media, and other sources. Based on the feedbacks automatically collected from users, city governments need to make decisions about which services to remove, offer, and improve. By using more digital tools and pioneering technologies, smart cities will provide smarter interactive services to users over the metaverse platform (Kohli et al., 2022). The environmental data (e.g., air quality, weather, energy consumption, traffic status, and available parking space) are fully replicated in the virtual world for user-friendly interface. Several smart services, such as utility payment and smart home control, can be now executed in the virtual world via the platforms and systems deployed in the metaverse: intelligent transportation systems (ITS), smart street light management systems, automatic parking systems, smart community portals, and indoor/outdoor video surveillance systems. Currently, the actual impact and benefit of these technologies to the smart cities is limited; however, the metaverse can be an accelerant to spread the presentation of smart services in the daily life of citizens (Mohammadi and Al-Fuqaha, 2018).

Among different technologies to enable smart cities in the physical world and in the metaverse, AI has shown a great significance to achieve automation and intelligence in smart services. In Park et al. (2019), a steady-state visual evoked potential-based BCI architecture was proposed by integrating EEG-based BCI, VR, and IoT technologies fueled by AI to collaboratively control home appliances. With the visual information captured through head-mounted display, the brain signals were recorded to analyze with ML algorithms and respond control commands with stimulation, which allow users to control home appliances over the IoT network. In Zhu et al. (2020a), an intelligent and ubiquitous IoT-enabled architecture was developed for to control and manage urban traffic, thus develop a hybrid ITS from the physical world to its virtual replication in the digital world

Many scenarios with practical data processing and decision making of different transportation services were investigated simultaneously in both of physical and virtual world, thus conducting high-quality real-time services to users and reducing operation and maintenance costs. Because of the hastiness of industrialization and the explosion of urbanization, air pollution has become a life-warning problem, which has affected living environment and physical health defectively. For early air pollution warning and management, the work (Du et al., 2021) proposed an efficient forecasting approach with a hybrid DL architecture, which combined 1-D CNNs and bi-directional LSTM networks to fully extract intrinsic correlation features and interdependence of multivariate time series data acquired from multiple sensors. Besides environmental pollution, sustainable agriculture has been attracting much more concerns in smart green cities (Bhat and Huang, 2021). In this context, AI is one of the vital information and communication (ICT) technologies, which has been widely used in precision agriculture systems with yield prediction, quality evaluation, and pest and disease detection.

Designing and implementing the metaverse ecosystem for smart cities with all administrative services, such as environment, education, transportation, culture, and other civil services, is really a challenging mission of metropolitan government. By gathering big data from multiple authenticated sources, many administrative services can be provided and improved in the metaverse thanks to AI technology for data analytics, in which usage rules, ethics, and security will be released to guarantee a safe experience environment.

3.4. Gaming

Gaming has always been a prime application in the metaverse, in which ML and DL are redefining and revolutionizing the gaming industry across multiple platforms, from console to mobile and PC platforms. This part will explore how ML and DL can revolutionize game development and how building a next gaming generation in the metaverse.

In the last decade, ML has had a huge impact on the way video games are developed. To build more realistic worlds with attractive challenges and unique stories, video game developers and studios have been increasingly turning to ML as a powerful tool set that help systems and NPCs to respond to player's action dynamically and reasonably. In Yannakakis and Togelius (2015), the role of artificial and computational intelligence in games has been discussed regarding many research topics: NPC behavior strategy and learning, tactical planning,

player response modeling, procedural content creation, player–NPC interaction design, general game AI, AI-assisted game story telling, and AI in commercial games. This work further investigated these topics according to three viewpoints: AI algorithms used in each topic, effectiveness of AI to human user in every topic, and human–computer interaction. For in-game decision making and learning, a comprehensive survey (Frutos-Pascual and Zapirain, 2017) investigated the use of AI algorithms in intelligent video and computer games. Regarding decision making, some primary AI algorithms (such as decision tree, fuzzy logic, Markov model, rule-based system, and finite-state machine) were deployed for different game developing tasks: modeling game flow, assessing playing motivation, evaluating immersive experience, adapting gameplay, adapting gaming strategy, customizing gameplay, and modeling and controlling NPC behavior. Besides, many learning-based tasks were accomplished with Naïve Bayes, ANN, SVM, and case-based reasoning system to classify user gameplay, classify NPC behavior, recognize user behavior, and adapt game flow based on personal experience.

In real-time strategy (RTS) games, such as StarCraft, Bayesian models have been used for modeling multi-scale uncertainty and multi-level abstraction levels: micromanagement, tactics, and strategy. These probabilistic learning models were able to cope with reactive units control, recognize objectives from tactical data, and predict opponent's gameplay based on strategic information (Synnaeve and Bessière, 2016). To reach an intelligent human-like response mechanism, several game software companies have applied AI in various testing tasks during the design and development stages. In Liaqat et al. (2020), a metamorphic testing mechanism was proposed to overcome the impracticability of controlling a large number of possible moving strategies in artificial chess games. This testing mechanism deployed a decision tree model to reveal metamorphic relations, which in turn effectively determine the optimal move among all possible ones. Based on the combination of RL and deep networks (Oh et al., 2021), AI agents were developed to address some inherent difficulties in real-time fighting games and to defeat pro players in one-vs-one battles. In addition to creating different fighting styles through a self-play curriculum, this deep RL framework was capable of all two-player competitive games which have level upgrade and balance policies. RL and supervised learning were also exploited to improve AI agents in RTS games (Barriga et al., 2019). Being superior to the Puppet search algorithm, CNNs and deep Q-learning networks inferred the outcome of the costly high-level search and optimized the available time to execute tactical searches. In a nutshell, AI with traditional ML and innovative DL algorithms has been making an unprecedented revolution in gaming experience in many aspects: improving the intelligence of NPCs, modeling complex systems, making games more beautiful and rational, conducting more realistic human-NPC interactions, reducing the cost of in-game world creation, and opening more opportunities of developing mobile games. We summarized the existing application-oriented AI-enabled works that showed the potential to be integrated and deployed in the metaverse in Table 2 along with a bibliometric analysis graph in Fig. 13.

3.5. Other potential applications

Besides healthcare, manufacturing, smart cities, and gaming, we have found some auxiliary business applications ahead for the metaverse.

E-commerce: For the purpose of which E-commerce has been integrated into the metaverse, numerous consumer brands have been diving into the digital world to create more delightful and seamless shopping experiences regardless of the unpopularity of VR devices for mainstream consumers. Many brands have moved forward step-by-step to build something entirely new by integrating digital stores which are able to bring the best offline and online shopping without any difference in user experience (Riedl, 2021). Indeed, virtual shopping can convey a remotely real-time experience of static products, in which

consumers, represented by an avatar, can walk around stores in a 3D rendered space and talk with a virtual cashier/seller powered by VR and AI technologies. Personalizing the customer experience is currently attracting more attention from retailers, not only for business survival but also for revenue growth, which can be performed effortlessly in the meta with AI-based shopping behavior understanding.

Human resources: Nowadays, many big tech companies are being creative to seek and communicate with young talents who are looking for jobs. The recruitment manners range from dispatching younger employees/leaders to online interview applicants with video calls to holding job fairs in the metaverse. Potential applicants can log in to the metaverse with a blockchain-aided authenticated account and then control their avatars to freely discuss with other avatars representing the company's human resource managers and project leaders (Handley, 2021). For recruitment guidance, the applicants can ask or receive help from a virtual assistant with AI-based NLP. In these kinds of recruitment events, the goal is to generate a friendly environment for both the recruiters and applicants for free-style communication, in which the applicants can actively discover more information about job positions instead of passively being asked questions by recruiters. In the last decade, emerging technologies (such as 5G, IoT, and DL) have brought workers/employees many convenient alternatives (fully remote and hybrid offline-online) to traditional work; however, the metaverse will revolutionize the future of work and the workplace. Recently, Horizon Workrooms (Meta, 2021) of Facebook, a well-designed meeting platform that allows users, represented as avatars, to work, collaborate, and communicate with others, besides training and coaching activities, in the virtual space by VR devices.

Real estate: We have seen huge investments from individual investors and institutions to virtual land in the metaverse. Some metaverses have already been released, including virtual gaming platforms like the Sandbox and Axie Infinity and virtual worlds like Decentraland and Upland, in which users can buy, sell, and trade things, including real estate (plots of land and virtual houses) (Kamin, 2021). These digital real estate, usually associated with non-fungible tokens (NFTs), are limited by supply to guarantee their values over time based on scarcity. The estate in the metaverse can be used as a virtual place for building constructions (houses and offices) or holding digital events (e.g., art exhibitions and fashion shows). Moreover, the metaverse is another cost-efficient channel for real estate companies to ultimately show the property to clients before making a decision. With a VR-aided immersive experience, the clients can discover the property, including interior and exterior from detailed furniture and overall structure, via VR tours and interactive walkthroughs.

Decentralized finance: Based on an open system of finance, decentralized finance (DeFi) is a cryptocurrency-based financial service that is regularly programmed through smart contracts to build exchanges besides providing many major services, such as lending, yield farming, and insurance without centralized authorities. Different from centralized finance which is controlled or managed by a centralized entity or a person, DeFi, with blockchain technology, facilitates financial services from peer-to-peer and allows users to fully control their assets while ensuring security and privacy. DeFi services are usually delivered via decentralized applications (Dapps) which are entirely built on open-sourced distributed platforms. By integrating DeFi (including basic and professional services) into the metaverse, users can make purchases of virtual products identified by NFTs in the digital world but will receive the real products in the real life. Furthermore, users can make profits in the metaverse based on the DeFi ecosystem with the lending, borrowing, mining, and staking of cryptocurrencies or other tokens. Users can provide liquidity to the liquidity pool with an underlying AI-based mechanism of a decentralized exchange to earn incentives. Swapping tokens (which can belong to the same chain or different chains) is the basic service that is prioritized to develop first on any Dapps.

Table 2
Summary of AI for the metaverse in the application aspect.

App. Aspect	Ref	Description	AI Technique
Healthcare	Hur et al. (2018)	Inertial sensory-based physical activity recognition for healthcare and wellness.	CNNs with an encoding algorithm to convert inertial sensory signals to color images.
	Huynh-The et al. (2021)	Physical activity recognition to support physicians and health-wellness experts in making decisions in living assistance and healthy risk awareness.	Residual CNNs to extract deep features in combination with handcrafted features based on an intermediate fusion mechanism.
	Qian et al. (2020)	Wearable devices based fall detection for IoT-based healthcare and medical services.	Hierarchical DL framework with CNNs for learning models at local devices and global cloud center.
	Ning et al. (2022)	Lesion segmentation in breast ultrasound images for detecting abnormalities.	Saliency-guided morphology-aware U-Net (SMU-Net) involving an additional middle feature learning stream and an auxiliary network.
	Qi et al. (2021)	Estimation of non-rigid heart motion in free-breathing 3D coronary magnetic resonance angiography images.	Deep encoder–decoder learning framework with CNN architectures.
Manufacturing	Doltsinis et al. (2018)	Automatic production system reconfigurations and upgrade to shorten product lifecycles and increase the number of product variants.	Symbiotic with reinforcement learning strategy by combining the learning capacity of Q-learning models and the domain knowledge of experts.
	Lee et al. (2021)	Reliable fault detection and diagnosis for quality inspection of production line in manufacturing.	Deep encoder–decoder framework using attentional RNNs.
	Xue et al. (2020)		Data-driven learning model with LSTM networks.
	Guo et al. (2019a)		Data-driven learning model using CNN-based transfer learning.
	Alkhalefah et al. (2021)	Forecasting the optimal buffer size required in production systems to maximize productivity.	A combination of a regular artificial neural network and a generic algorithm.
	Huang et al. (2019)	Production progress prediction to optimize on-time order delivery activities in manufacturing enterprises.	Data-driven learning model using two-stage transfer learning mechanism on deep belief network architecture.
Smart Cities	Park et al. (2019)	Smart home appliances control and management over IoT networks.	A set of AI algorithms for brain signal processing and decision making to respond stimulation commands.
	Zhu et al. (2020a)	Intelligent and ubiquitous IoT-enabled urban traffic control and management in ITS systems.	Hierarchical AI framework by deploying ML algorithms in the digital world to automatically make decisions in the physical world.
	Du et al. (2021)	Early air pollution warning and management in smart environment surveillance systems.	DL framework with 1-D CNNs coupled with bi-directional LSTM networks.
	Bhat and Huang (2021)	Sustainable agriculture from smart farm to intelligent food production line.	AI framework for supervised learning with conventional ML algorithms applied in detection, classification, and recognition models.
Gaming	Synnaeve and Bessière (2016)	Modeling multi-scale uncertainty and multi-level abstraction levels in RTS games.	Multi-scale Bayesian models with probabilistic learning for multi-level units control, tactics, and strategy.
	Liaqat et al. (2020)	Design of a metamorphic testing mechanism for game flow evaluation in artificial chess games.	ML framework with decision tree algorithm to determine the optimal move among all possible ones for AI agents.
	Oh et al. (2021)	Development of AI agents in player-vs-NPC and player-vs-player fighting games.	A combination of RL and deep networks with different architectures, such as RNN and CNN.
	Barriga et al. (2019)	Development of AI agents and improvement of tactical intelligence in RTS games.	Supervised learning framework with CNN architectures and reinforcement learning with deep Q-learning networks.

4. Metaverse projects

This section briefly introduces some attractive metaverse projects, including Decentraland, Sandbox, Realy, Star Atlas, Bit.Country, and DeHealth, which have applied AI to deliver multifarious services and applications in the virtual world, from real estate to E-commerce and real estate. The landscapes inside the virtual worlds of the projects are shown in Fig. 14, for DeHealth, virtual doctors as avatars in the metaverse.

Decentraland[3]: This is a decentralized virtual reality platform built on the Ethereum blockchain, in which users can experience, create, and monetize assets, content, and applications. In Decentraland, virtual land is determined as a non-fungible, transferable, and scarce digital asset recorded by the Ethereum smart contract. Different from traditional virtual worlds and social networks, Decentraland is not controlled by any centralized organization; that is, no single agent has permission to modify the rules of software, content, and economics of cryptocurrency, or prevent others from accessing the world, trading digital products, and experienced services. The traversable 3D world in Decentraland allows embedding immersive components and adjacency to create content that makes this project attractive and unique. A scripting programming language of Decentraland enables developers to easily code AI-based service-oriented applications for users and accordingly encourage users to create new content. Some principal use cases are content curation, advertising, digital collectibles, and social besides other minor ones, such as education, virtual tourism, healthcare, and virtual shopping. Regarding architecture, the Decentraland protocol has three layers: a consensus layer to track land ownership and its content, a land content layer to distribute the materials for rendering via a decentralized storage system, and a real-time layer to establish peer-to-peer connections for world view. The native token, for in-world purchase goods and services, of Decentraland is MANA, a fungible token built on the ERC-20 (Ethereum Request for Comments 20) protocol.
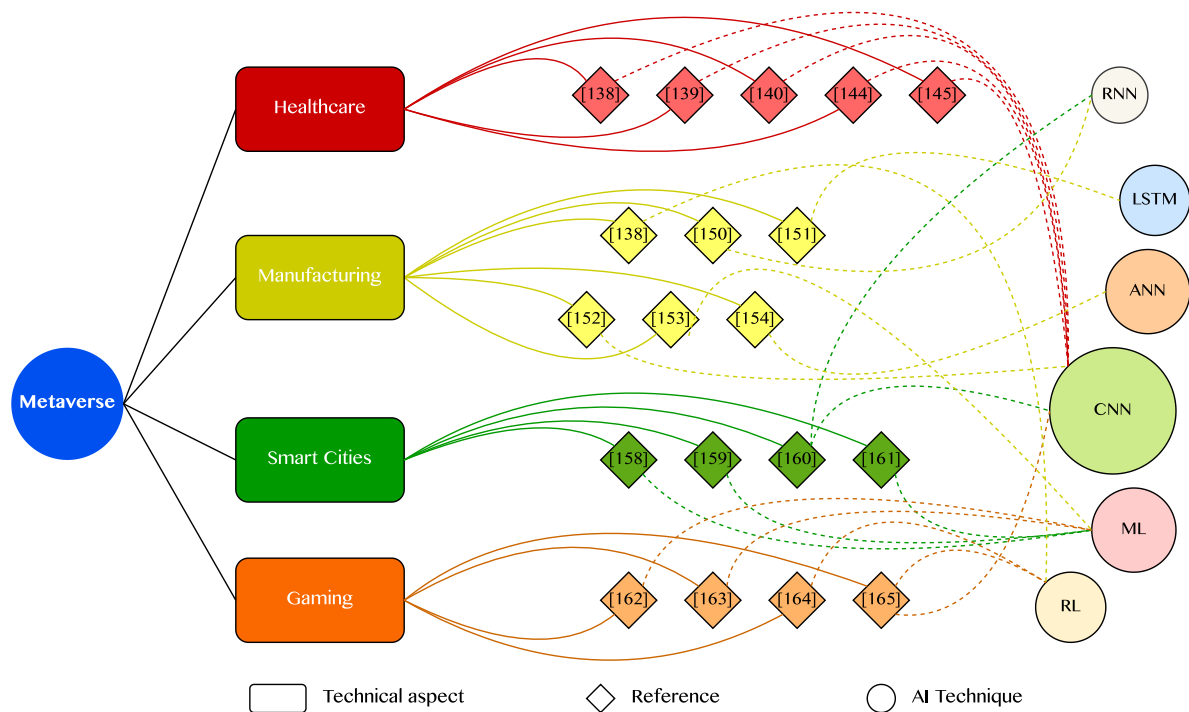
---

[3] https://decentraland.org/

Fig. 13. A bibliometric analysis and visualization of potential AI-based application works for the metaverse.

Fig. 14. Inside the virtual worlds of different metaverse projects (left to right): Decentraland, Sandbox, Realy, Star Atlas, Bit.Country, and DeHealth.

Sandbox[4]: The Sandbox metaverse is a user-generated decentralized Ethereum blockchain-based virtual world, which allows users and gamers to build, own, and monetize gaming experiences. Inspired by Minecraft,[5] the Sandbox metaverse is first built as a 2D mobile pixel game and then extended to a fully-fledged 3D world with a voxel gaming platform, wherein users are capable of playing, sharing, collecting, and trading virtual goods and services without central control. Remarkably, creators can earn SAND, the native token of Sandbox, by selling their creations on a marketplace with secure copyright ownership which is associated and guaranteed via NFT, i.e., every item in the metaverse will be authenticated by a unique and immutable blockchain mechanism. As the primary use cases of SAND in the Sandbox metaverse, the token holders can access and experience the virtual world, vote governance decisions via DAO mechanism, stake tokens to earn revenues, and donate tokens as incentives to developers for the metaverse growth. Besides the blockchain technology with ERC-20 to generate SAND tokens and ERC-1155 for digital assets trading, AI has been utilized in the Sandbox metaverse. For example, as a powerful toolkit, gaming coders can deploy ML models to improve the intelligence of virtual agents/assistants and DL models to enhance the render quality, and developers can leverage different AI frameworks to minimize gaming crashes and errors. Building and training ML/DL models are trouble-free with intuitive high-level APIs.

Realy[6]: From the real world to the fully virtual world by creating a unique metaverse ecological world, the Realy metaverse is defined as a super-realistic, futuristic, technologically conscious world, in which

E-commerce, social, gaming, and trading are truly integrated to bring a seamless virtual-reality experience to users. In the virtual world of Realy metaverse, users can enjoy colorful journeys through their personalized avatars with 3D virtual clothes which are available in the marketplace and linked with unique NFTs. There is an interesting feature of Realy compared with other metaverse projects, which is about avatar control and management procedures. When users are online, their avatars bring an immersive experience. When users are offline, the avatars are driven by a set of AI-based self-discipline systems. The whole virtual world is automatically operated, controlled, and managed by a decentralized DAO organization. Regarding technology in the Realy metaverse, besides VR and blockchain, AI is adopted in many aspects to generally improve the user immersive experience, such as enhancing 3D visual rendered effects, boosting the intelligence of avatars for realistic behaviors, and integrating VR and holographic projection.

Star Atlas[7]: As one of the most recent innovative metaverse projects, Star Atlas introduces a virtual gaming world built on the integration of multiplayer video game platforms, real-time immersive experiences with 3D rendered visualization, blockchain-based decentralized financial services, and AI-powered game engines. Star Atlas helps to complete an ecosystem on the Solana blockchain by filling the gap between metaverse and blockchain technology. In the gaming metaverse of Star Atlas, users can trade digital assets like land, equipment, crew, ship, and components by using the in-game cryptocurrency token known as POLIS which can be used for multiple cross-metaverse games. To tune the gameplay more logical and realistic, ML algorithms are applied to

---

[4] https://www.sandbox.game/en/
[5] https://www.minecraft.net/en-us
[6] https://realy.pro/
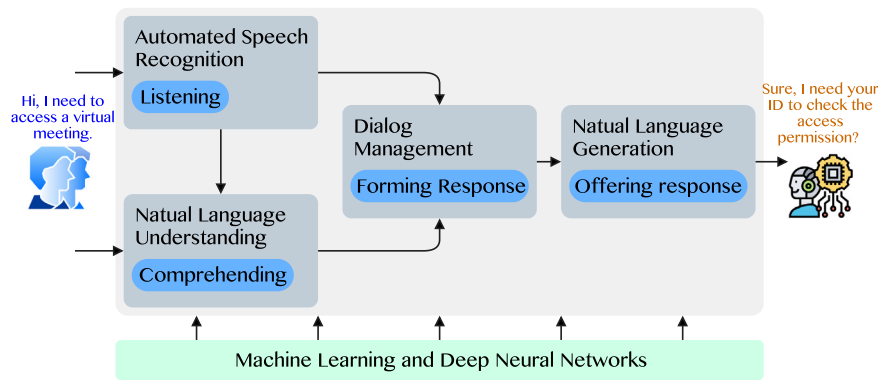
[7] https://staratlas.com/

Fig. 15. General processing flow of conversational AI to deliver contextual and personal experience to users.

improve the intelligence of NPCs and AI agents in the tactical planning of actions and fighting strategy in player–NPC combats.

Bit.Country[8]: As a user-oriented metaverse project, Bit.Country builds a 3D virtual world for everyone who can set up its own community in the metaverse with rules and operations to attract followers and contributors. By introducing a new level of virtual social interaction, Bit.Country has two platforms: one is a traditional web view for content creation and service provision, and another is an addition 3D gaming view for VR-aided immersive experience. In each individual community configured by users holding platform tokens, all the rules are managed and linked together by AI models to ensure logical operations without any conflicts. Being more than a virtual world, Bit.Country is capable of connecting some virtual aspects to the real world to obtain a sustainable future instead of a joyful immersive experience.

DeHealth[9]: Introduced by a British non-profit organization, DeHealth is the world's first decentralized healthcare metaverse, which allows doctors and patients to work and interact with each other in a full 3D virtual world. In the DeHealth metaverse, several high-quality healthcare and medical services are delivered, such as health analytics on the go, recommendations from advanced AI-bot, and real-time dialogues with doctors and health experts worldwide. To encourage informative data sharing activities, the metaverse has some trading cryptocurrency pools for users and patients to earn assets by selling anonymized medical data. The data collected from a decentralized network will be used to build AI-based diagnostic models for diversified tasks in the healthcare and medical domains. In the metaverse, doctors and patients are able to communicate via a virtual space replicating the real-world environment. Some virtual hospitals and healthcare centers can be built to provide virtual services with real data and diagnosis results. The DeHealth metaverse is in consideration to be extended with additional education-oriented services based on VR technology.

5. Conclusion and research directions

In this survey, we have comprehensively investigated the role of AI in the foundation of the metaverse and its potential to enhance user immersive experience in the virtual world. At the beginning of this work, the fundamental concepts of the metaverse and AI techniques have been provided, along with the role of AI in the metaverse. Subsequently, several principal technical aspects, such as NLP, machine vision, blockchain, networking, DT, and neural interface, and many application aspects, such as healthcare, manufacturing, smart cities, gaming, E-commerce, and DeFi, have been analyzed. The reviewed AI-based solutions have shown that AI has great potential to toughen the systems infrastructure, uplift the 3D immersive experience, and flourish the built-in services in the virtual worlds significantly. Finally, we have

examined prominent metaverse projects, in which AI techniques were used to sharpen the quality of services and encompass the ecosystem of the metaverse.

We now delineate some AI research directions in the metaverse. Being more advanced than regular virtual personal assistants which are developed for a general purpose with simple dialog management, virtual customers/employee assistants powered by conversational AI can serve many specific purposes of multi-level philosophical conversations to enhance user interactive experience. Conversational AI with a processing flow in Fig. 15 is a set of technologies (e.g., automatic speech recognition, language processing, advanced dialog management, and ML) that can offer human-like interactions in the metaverse based on recognizing speech and text, understanding intention, deciphering various languages, and responding human-mimicking conversations over voice modality.

Most of the current metaverse projects limit users to explore, own and customize things in the virtual world. In the future, users will be allowed to create hyperreal objects and content easily and quickly with the help of AI. Various kinds of hyperreal objects (e.g., faces, bodies, plants, animals, vehicles, buildings, and other inanimate objects) can be remixed endlessly by users to make unique experiences and excite creation. Accordingly, the combination of VR and AI-based content generation can bring a complete immersion in alternative realities. In this context, AI tools should be cheap to everyone and have user-friendly interfaces. Further, ethical issues relating to user-generated metaverse need to be seriously examined with constraints and policies between users and third-party organizations to mitigate risks and harmful threats to individuals and societies when users synthesize hyperreal media content.

In many AI-aided services and applications in the metaverse, the decisions are made by AI agents, which are driven by ML models as black boxes without the capability of interpretability and explainability. Metaverse developers, virtual world designers, and users cannot completely understand AI decision-making processes (e.g., how and why an AI model delivers a prediction), and probably trust them blindly. To overcome these problems, explainable AI (XAI) is a set of tools and methods to describe AI models, analyze their expected impacts, characterize model transparency, and examine outcomes, allowing human users to entirely comprehend and trust the AI models with end-to-end process monitoring and accountability. With XAI, system engineers and data scientists who apply AI in the metaverse (from system infrastructure to services and applications in the virtual worlds) can understand and explain what exactly is happening inside an AI model, how is a specific result generated by an AI algorithm, and when is a prediction model likely to crash. Besides increasing end-user confidence, model auditability, and operative efficiency, XAI mitigates legal risks and security threats of production AI in the metaverse while guaranteeing users' reliable experience.

---

[8] https://bit.country/
[9] https://www.dehealth.world/

CRediT authorship contribution statement

Thien Huynh-The: Conceptualization, Methodology, Investigation, Writing – original draft, Visualization. **Quoc-Viet Pham:** Writing – original draft, Conceptualization, Writing – review & editing. **Xuan-Qui Pham:** Writing – original draft, Writing – review & editing, Visualization. **Thanh Thi Nguyen:** Formal analysis, Investigation, Formal analysis. **Zhu Han:** Supervision, Writing – review & editing, Resources. **Dong-Seong Kim:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Thien Huynh-The reports financial support was provided by National Research Foundation of Korea. Dong-Seong Kim reports a relationship with National Research Foundation of Korea that includes: funding grants.

Data availability

No data was used for the research described in the article.

References

Abibullaev, B., Zollanvari, A., 2019. Learning discriminative spatiospectral features of ERPs for accurate brain–computer interfaces. IEEE J. Biomed. Health Inf. 23 (5), 2009–2020.

Agushaka, J.O., Ezugwu, A.E., Abualigah, L., 2022. Dwarf mongoose optimization algorithm. Comput. Methods Appl. Mech. Engrg. 391, 114570.

Alkhalefah, H., Qudeiri, J.E.A., Umer, U., Abidi, M.H., Elkaseer, A., 2021. Development of an efficient prediction model for optimal design of serial production lines. IEEE Access 9, 61807–61818.

Alsenwi, M., Tran, N.H., Bennis, M., Pandey, S.R., Bairagi, A.K., Hong, C.S., 2021. Intelligent resource slicing for eMBB and URLLC coexistence in 5G and beyond: A deep reinforcement learning based approach. IEEE Trans. Wireless Commun. 20 (7), 4585–4600.

Athiwaratkun, B., Stokes, J.W., 2017. Malware classification with LSTM and GRU language models and a character-level CNN. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, New Orleans, LA, USA, pp. 2482–2486.

Azamfar, M., Li, X., Lee, J., 2020. Deep learning-based domain adaptation method for fault diagnosis in semiconductor manufacturing. IEEE Trans. Semicond. Manuf. 33 (3), 445–453.

Azari, A., Ozger, M., Cavdar, C., 2019. Risk-aware resource allocation for URLLC: Challenges and strategies with machine learning. IEEE Commun. Mag. 57 (3), 42–48.

Banos, O., Bang, J., Hur, T., Siddiqi, M.H., Thien, H.-T., Vui, L.-B., Ali Khan, W., Ali, T., Villalonga, C., Lee, S., 2015. Mining human behavior for health promotion. In: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. EMBC, IEEE, Milan, Italy, pp. 5062–5065.

Barba, E., MacIntyre, B., Mynatt, E.D., 2012. Here we are! Where are we? Locating mixed reality in the age of the smartphone. Proc. IEEE 100 (4), 929–936.

Barriga, N.A., Stanescu, M., Besoain, F., Buro, M., 2019. Improving RTS game AI by supervised policy learning, tactical search, and deep reinforcement learning. IEEE Comput. Intell. Mag. 14 (3), 8–18.

Benes, K., Baskar, M.K., Burget, L., 2017. Residual memory networks in language modeling: Improving the reputation of feed-forward networks. In: Proceedings of the INTERSPEECH. INTERSPEECH, Stockholm, Sweden, pp. 284–288.

Bernal, S.L., Celdrán, A.H., Pérez, G.M., Barros, M.T., Balasubramaniam, S., 2022. Security in brain-computer interfaces: State-of-the-art, opportunities, and future challenges. ACM Comput. Surv. 54 (1), 1–35.

Bhat, S.A., Huang, N.-F., 2021. Big data and AI revolution in precision agriculture: Survey and challenges. IEEE Access 9, 110209–110222.

Boopalan, P., Ramu, S.P., Pham, Q.-V., Dev, K., Maddikunta, P.K.R., Gadekallu, T.R., Huynh-The, T., et al., 2022. Fusion of federated learning and industrial internet of things: A survey. Comput. Netw. 109048.

Cannavò, A., Lamberti, F., 2021. How blockchain, virtual reality, and augmented reality are converging, and why. IEEE Consum. Electron. Mag. 10 (5), 6–13.

Chen, M., Challita, U., Saad, W., Yin, C., Debbah, M., 2019. Artificial neural networks-based machine learning for wireless networks: A tutorial. IEEE Commun. Surv. Tutor. 21 (4), 3039–3071.

Chen, K., Gong, S., Xiang, T., 2011. Human pose estimation using structural support vector machines. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. ICCV Workshops, IEEE, Barcelona, Spain, pp. 846–851.

Chen, X., Kundu, K., Zhu, Y., Ma, H., Fidler, S., Urtasun, R., 2018a. 3D object proposals using stereo imagery for accurate object class detection. IEEE Trans. Pattern Anal. Mach. Intell. 40 (5), 1259–1272.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018b. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848.

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.

Cheng, K.-W., Chen, Y.-T., Fang, W.-H., 2018. Improved object detection with iterative localization refinement in convolutional neural networks. IEEE Trans. Circuits Syst. Video Technol. 28 (9), 2261–2275.

Dang, Q., Yin, J., Wang, B., Zheng, W., 2019. Deep learning based 2D human pose estimation: A survey. Tsinghua Sci. Technol. 24 (6), 663–676.

Daniluk, M., Rocktäschel, T., Welbl, J., Riedel, S., 2017. Frustratingly short attention spans in neural language modeling. arXiv preprint arXiv:1702.04521.

Darvishi, H., Ciuonzo, D., Eide, E.R., Rossi, P.S., 2021. Sensor-fault detection, isolation and accommodation for digital twins via modular data-driven architecture. IEEE Sens. J. 21 (4), 4827–4838.

Doltsinis, S., Ferreira, P., Lohse, N., 2018. A symbiotic human–machine learning approach for production ramp-up. IEEE Trans. Hum.-Mach. Syst. 48 (3), 229–240.

Dong, C., Loy, C.C., He, K., Tang, X., 2016. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 38 (2), 295–307.

Dong, R., She, C., Hardjawana, W., Li, Y., Vucetic, B., 2019. Deep learning for hybrid 5G services in mobile edge computing systems: Learn from a digital twin. IEEE Trans. Wireless Commun. 18 (10), 4692–4707.

Du, S., Li, T., Yang, Y., Horng, S.-J., 2021. Deep air quality forecasting using hybrid deep learning framework. IEEE Trans. Knowl. Data Eng. 33 (6), 2412–2424.

Elayan, H., Aloqaily, M., Guizani, M., 2021. Digital twin for intelligent context-aware IoT healthcare systems. IEEE Internet Things J. 8 (23), 16749–16757.

Ezugwu, A.E., Agushaka, J.O., Abualigah, L., Mirjalili, S., Gandomi, A.H., 2022. Prairie dog optimization algorithm. Neural Comput. Appl. 1–49.

Fan, S., Zhang, H., Zeng, Y., Cai, W., 2021. Hybrid blockchain-based resource trading system for federated learning in edge computing. IEEE Internet Things J. 8 (4), 2252–2264.

Feng, M., Gilani, S.Z., Wang, Y., Zhang, L., Mian, A., 2021. Relation graph network for 3D object detection in point clouds. IEEE Trans. Image Process. 30, 92–107.

Frutos-Pascual, M., Zapirain, B.G., 2017. Review of the use of AI techniques in serious games: Decision making and machine learning. IEEE Trans. Comput. Intell. AI Games 9 (2), 133–152.

Gadekallu, T.R., Pham, Q.-V., Nguyen, D.C., Maddikunta, P.K.R., Deepa, N., Prabadevi, B., Pathirana, P.N., Zhao, J., Hwang, W.-J., 2022. Blockchain for edge of things: applications, opportunities, and challenges. IEEE Internet Things J. 9 (2), 964–988.

Ghandar, A., Ahmed, A., Zulfiqar, S., Hua, Z., Hanai, M., Theodoropoulos, G., 2021. A decision support system for urban agriculture using digital twin: A case study with aquaponics. IEEE Access 9, 35691–35708.

Gu, B., Zhang, X., Lin, Z., Alazab, M., 2021. Deep multiagent reinforcement-learning-based resource allocation for internet of controllable things. IEEE Internet Things J. 8 (5), 3066–3074.

Guo, L., Lei, Y., Xing, S., Yan, T., Li, N., 2019a. Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. IEEE Trans. Ind. Electron. 66 (9), 7316–7325.

Guo, S., Lin, Y., Li, S., Chen, Z., Wan, H., 2019b. Deep spatial–temporal 3D convolutional neural networks for traffic data forecasting. IEEE Trans. Intell. Transp. Syst. 20 (10), 3913–3926.

Hackl, C., 2021. Defining the metaverse today.

Han, Y., Yu, C., Li, D., Zhang, J., Lai, Y., 2020. Accuracy analysis on 360° virtual reality video quality assessment methods. In: Proceedings of the IEEE/ACM 13th International Conference on Utility and Cloud Computing. UCC, IEEE, Leicester, UK, pp. 414–419.

Handley, L., 2021. Looking for a job? You might get hired via the metaverse, experts say. URL https://www.cnbc.com/2021/11/30/looking-for-a-job-you-might-get-hired-via-the-metaverse-experts-say.html.

Hassanzadeh, T., Essam, D., Sarker, R., 2021. 2D to 3D evolutionary deep convolutional neural networks for medical image segmentation. IEEE Trans. Med. Imaging 40 (2), 712–721.

He, H., Liu, G., Zhu, X., He, L., Tian, G., 2019. Interacting multiple model-based human pose estimation using a distributed 3D camera network. IEEE Sens. J. 19 (22), 10584–10590.

He, H., Wu, D., 2020. Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. IEEE Trans. Biomed. Eng. 67 (2), 399–410.

Ho, C.D., Nguyen, T.-V., Huynh-The, T., Nguyen, T.-T., da Costa, D.B., An, B., 2021. Short-packet communications in wireless-powered cognitive IoT networks: Performance analysis and deep learning evaluation. IEEE Trans. Veh. Technol. 70 (3), 2894–2899.

Hu, Z., Bulling, A., Li, S., Wang, G., 2021. Fixationnet: Forecasting eye fixations in task-oriented virtual environments. IEEE Trans. Vis. Comput. Graphics 27 (5), 2681–2690.

Hu, Z., Li, S., Zhang, C., Yi, K., Wang, G., Manocha, D., 2020. DGaze: CNN-based gaze prediction in dynamic scenes. IEEE Trans. Vis. Comput. Graphics 26 (5), 1902–1911.

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 J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (5), 2015–2030.

Hua, C.-H., Huynh-The, T., Bae, S.-H., Lee, S., 2020a. Cross-attentional bracket-shaped convolutional network for semantic image segmentation. Inform. Sci. 539, 277–294.

Hua, C.-H., Huynh-The, T., Kim, K., Yu, S.-Y., Le-Tien, T., Park, G.H., Bang, J., Khan, W.A., Bae, S.-H., Lee, S., 2019. Bimodal learning via trilogy of skip-connection deep networks for diabetic retinopathy risk progression identification. Int. J. Med. Inform. 132, 103926.

Hua, C.-H., Huynh-The, T., Lee, S., 2018. Convolutional networks with bracket-style decoder for semantic scene segmentation. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. SMC, IEEE, Miyazaki, Japan, pp. 2980–2985.

Hua, C.-H., Huynh-The, T., Lee, S., 2020b. DRAN: Densely reversed attention based convolutional network for diabetic retinopathy detection. In: Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. EMBC, IEEE, Montreal, QC, Canada, pp. 1992–1995.

Hua, C.-H., Kim, K., Huynh-The, T., You, J.I., Yu, S.-Y., Le-Tien, T., Bae, S.-H., Lee, S., 2021. Convolutional network with twofold feature augmentation for diabetic retinopathy recognition from multi-modal images. IEEE J. Biomed. Health Inf. 25 (7), 2686–2697.

Huang, S., Guo, Y., Liu, D., Zha, S., Fang, W., 2019. A two-stage transfer learning-based deep learning approach for production progress prediction in IoT-enabled manufacturing. IEEE Internet Things J. 6 (6), 10627–10638.

Hur, T., Bang, J., Lee, J., Kim, J.-I., Lee, S., et al., 2018. Iss2Image: A novel signal-encoding technique for CNN-based human activity recognition. Sensors 18 (11), 3910.

Huynh-The, T., Banos, O., Lee, S., Kang, B.H., Kim, E.-S., Le-Tien, T., 2017. NIC: A robust background extraction algorithm for foreground detection in dynamic scenes. IEEE Trans. Circuits Syst. Video Technol. 27 (7), 1478–1490.

Huynh-The, T., Hua, C.-H., Kim, D.-S., 2020a. Encoding pose features to images with data augmentation for 3-D action recognition. IEEE Trans. Ind. Inf. 16 (5), 3100–3111.

Huynh-The, T., Hua, C.-H., Pham, Q.-V., Kim, D.-S., 2020b. MCNet: An efficient CNN architecture for robust automatic modulation classification. IEEE Commun. Lett. 24 (4), 811–815.

Huynh-The, T., Hua, C.-H., Tu, N.A., Hur, T., Bang, J., Kim, D., Amin, M.B., Kang, B.H., Seung, H., Shin, S.-Y., et al., 2018. Hierarchical topic modeling with pose-transition feature for action recognition using 3D skeleton data. Inform. Sci. 444, 20–35.

Huynh-The, T., Hua, C.-H., Tu, N.A., Kim, D.-S., 2020c. Learning 3D spatiotemporal gait feature by convolutional network for person identification. Neurocomputing 397, 192–202.

Huynh-The, T., Hua, C.-H., Tu, N.A., Kim, D.-S., 2021. Physical activity recognition with statistical-deep fusion model using multiple sensory data for smart health. IEEE Internet Things J. 8 (3), 1533–1543.

Huynh-The, T., Le, B.-V., Lee, S., Le-Tien, T., Yoon, Y., et al., 2014. Using weighted dynamic range for histogram equalization to improve the image contrast. EURASIP J. Image Video Process. 2014 (1), 1–17.

Huynh-The, T., Le, B.-V., Lee, S., Yoon, Y., 2016. Interactive activity recognition using pose-based spatio–temporal relation features and four-level Pachinko Allocation Model. Inform. Sci. 369, 317–333.

Jeong, J.-H., Shim, K.-H., Kim, D.-J., Lee, S.-W., 2020. Brain-controlled robotic arm system based on multi-directional CNN-BiLSTM network using EEG signals. IEEE Trans. Neural Syst. Rehabil. Eng. 28 (5), 1226–1238.

Jin, Y., Chen, M., Goodall, T., Patney, A., Bovik, A.C., 2021. Subjective and objective quality assessment of 2D and 3D foveated video compression in virtual reality. IEEE Trans. Image Process. 30, 5905–5919.

Jin, Z., Iqbal, M.Z., Bobkov, D., Zou, W., Li, X., Steinbach, E., 2020a. A flexible deep CNN framework for image restoration. IEEE Trans. Multimed. 22 (4), 1055–1068.

Jin, N., Wu, J., Ma, X., Yan, K., Mo, Y., 2020b. Multi-task learning model based on multi-scale CNN and LSTM for sentiment classification. IEEE Access 8, 77060–77072.

Kamin, D., 2021. Investors snap up metaverse real estate in a virtual land boom. URL https://www.nytimes.com/2021/11/30/business/metaverse-real-estate.html.

Kanterman, M., Naidu, N., 2021. Metaverse may be $800 billion market, next tech platform. URL https://www.bloomberg.com/professional/blog/metaverse-may-be-800-billion-market-next-tech-platform/.

Kataoka, K., Yamamoto, T., Otsuki, M., Shibata, F., Kimura, A., 2019. A new interactive haptic device for getting physical contact feeling of virtual objects. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces. VR, IEEE, Osaka, Japan, pp. 1323–1324.

Kersten-Oertel, M., Jannin, P., Collins, D.L., 2012. DVV: a taxonomy for mixed reality visualization in image guided surgery. IEEE Trans. Vis. Comput. Graphics 18 (2), 332–352.

Khan, M.A., Abbas, S., Rehman, A., Saeed, Y., Zeb, A., Uddin, M.I., Nasser, N., Ali, A., 2021. A machine learning approach for blockchain-based smart home networks security. IEEE Netw. 35 (3), 223–229.

Khysru, K., Jin, D., Huang, Y., Feng, H., Dang, J., 2021. A Tibetan language model that considers the relationship between suffixes and functional words. IEEE Signal Process. Lett. 28, 459–463.

Kiumarsi, B., Vamvoudakis, K.G., Modares, H., Lewis, F.L., 2018. Optimal and autonomous control using reinforcement learning: A survey. IEEE Trans. Neural Netw. Learn. Syst. 29 (6), 2042–2062.

Kohli, V., Tripathi, U., Chamola, V., Rout, B.K., Kanhere, S.S., 2022. A review on Virtual Reality and Augmented Reality use-cases of Brain Computer Interface based applications for smart cities. Microprocess. Microsyst. 88, 104392.

Kotsiantis, S.B., 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.

Kumar, P., Kumar, R., Srivastava, G., Gupta, G.P., Tripathi, R., Gadekallu, T.R., Xiong, N., 2021. PPSF: A privacy-preserving and secure framework using blockchain-based machine-learning for IoT-driven smart cities. IEEE Trans. Netw. Sci. Eng. 8 (3), 2326–2341.

Lahiri, A., Bairagya, S., Bera, S., Haldar, S., Biswas, P.K., 2020. Lightweight modules for efficient deep learning based image restoration. IEEE Trans. Circuits Syst. Video Technol. 31 (4), 1395–1410.

Lai, P.K., Xie, S., Lang, J., Laganière, R., 2019. Real-time panoramic depth maps from omni-directional stereo images for 6 DoF videos in virtual reality. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces. VR, IEEE, Osaka, Japan, pp. 405–412.

Lee, H., Jeong, H., Koo, G., Ban, J., Kim, S.W., 2021. Attention recurrent neural network-based severity estimation method for interturn short-circuit fault in permanent magnet synchronous machines. IEEE Trans. Ind. Electron. 68 (4), 3445–3453.

Lee, J., Lee, J., Yoo, H.-J., 2020. SRNPU: An energy-efficient CNN-based super-resolution processor with tile-based selective super-resolution in mobile devices. IEEE J. Emerg. Sel. Top. Circuits Syst. 10 (3), 320–334.

Li, J., Shao, Y., Wei, K., Ding, M., Ma, C., Shi, L., Han, Z., Poor, H.V., 2021. Blockchain assisted decentralized federated learning (BLADE-FL): Performance analysis and resource allocation. arXiv preprint arXiv:2101.06905.

Li, H., Shrestha, A., Heidari, H., Le Kernec, J., Fioranelli, F., 2020. Bi-LSTM network for multimodal continuous human activity recognition and fall detection. IEEE Sens. J. 20 (3), 1191–1201.

Liaqat, A., Sindhu, M.A., Siddiqui, G.F., 2020. Metamorphic testing of an artificially intelligent chess game. IEEE Access 8, 174179–174190.

Lins, R.G., Givigi, S.N., 2021. Cooperative robotics and machine learning for smart manufacturing: Platform design and trends within the context of industrial internet of things. IEEE Access 9, 95444–95455.

Liu, D., Bober, M., Kittler, J., 2021a. Visual semantic information pursuit: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 43 (4), 1404–1422.

Liu, Y., Chen, K., Liu, C., Qin, Z., Luo, Z., Wang, J., 2019a. Structured knowledge distillation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Long Beach, CA, USA, pp. 2604–2613.

Liu, D., Fu, J., Qu, Q., Lv, J., 2019b. BFGAN: backward and forward generative adversarial networks for lexically constrained sentence generation. IEEE/ACM Trans. Audio Speech Language Process. 27 (12), 2350–2361.

Liu, N., Han, J., Yang, M.-H., 2020a. Picanet: Pixel-wise contextual attention learning for accurate saliency detection. IEEE Trans. Image Process. 29, 6438–6451.

Liu, B., Yin, G., 2020. Chinese document classification with bi-directional convolutional language model. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York, NY, United States, pp. 1785–1788.

Liu, Y., Yu, F.R., Li, X., Ji, H., Leung, V.C.M., 2020b. Blockchain and machine learning for communications and networking systems. IEEE Commun. Surv. Tutor. 22 (2), 1392–1431.

Liu, H., Zhang, S., Zhang, P., Zhou, X., Shao, X., Pu, G., Zhang, Y., 2021b. Blockchain and federated learning for collaborative intrusion detection in vehicular edge computing. IEEE Trans. Veh. Technol. 70 (6), 6073–6084.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Boston, MA, USA, pp. 3431–3440.

Lu, Y., Huang, X., Dai, Y., Maharjan, S., Zhang, Y., 2020. Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. IEEE Trans. Ind. Inf. 16 (6), 4177–4186.

Luo, C., Ji, J., Wang, Q., Chen, X., Li, P., 2020. Channel state information prediction for 5G wireless communications: A deep learning approach. IEEE Trans. Netw. Sci. Eng. 7 (1), 227–236.

Lv, Z., Qiao, L., Wang, Q., Piccialli, F., 2021. Advanced machine-learning methods for brain-computer interfacing. IEEE/ACM Trans. Comput. Biol. Bioinform. 18 (5), 1688–1698.

Ma, W., Cui, Y., Si, C., Liu, T., Wang, S., Hu, G., 2020. CharBERT: Character-aware pre-trained language model. arXiv preprint arXiv:2011.01513.

Ma, R., Yu, T., Zhong, X., Yu, Z.L., Li, Y., Gu, Z., 2021. Capsule network for ERP detection in brain-computer interface. IEEE Trans. Neural Syst. Rehabil. Eng. 29, 718–730.

Maddikunta, P.K.R., Pham, Q.-V., Nguyen, D.C., Huynh-The, T., Aouedi, O., Yenduri, G., Bhattacharya, S., Gadekallu, T.R., 2022. Incentive techniques for the internet of things: a survey. J. Netw. Comput. Appl. 103464.

Majeed, U., Khan, L.U., Yousafzai, A., Han, Z., Park, B.J., Hong, C.S., 2021. ST-BFL: A structured transparency empowered cross-silo federated learning on the blockchain framework. IEEE Access 9, 155634–155650.

Matran-Fernandez, A., Poli, R., 2017. Brain–computer interfaces for detection and localization of targets in aerial images. IEEE Trans. Biomed. Eng. 64 (4), 959–969.

Mei, S., Jiang, R., Li, X., Du, Q., 2020. Spatial and spectral joint super-resolution using convolutional neural network. IEEE Trans. Geosci. Remote Sens. 58 (7), 4590–4603.

Meta, 2021. Introducing horizon workrooms: Remote collaboration reimagined. URL https://about.fb.com/news/2021/08/introducing-horizon-workrooms-remote-collaboration-reimagined/.

Minaee, S., Boykov, Y.Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N., Terzopoulos, D., 2021. Image segmentation using deep learning: A survey. IEEE Trans. Pattern Anal. Mach. Intell. (Early Access).

Mohammadi, M., Al-Fuqaha, A., 2018. Enabling cognitive smart cities using big data and machine learning: Approaches and challenges. IEEE Commun. Mag. 56 (2), 94–101.

Moore, S.K., 2022. Meta aims to build the world's fastest AI supercomputer. URL https://spectrum.ieee.org/meta-ai-supercomputer.

Nguyen, D.C., Pathirana, P.N., Ding, M., Seneviratne, A., 2020. Privacy-preserved task offloading in mobile blockchain with deep reinforcement learning. IEEE Trans. Netw. Serv. Manag. 17 (4), 2536–2549.

Ning, Z., Zhong, S., Feng, Q., Chen, W., Zhang, Y., 2022. SMU-Net: Saliency-guided morphology-aware U-Net for breast lesion segmentation in ultrasound image. IEEE Trans. Med. Imaging 41 (2), 476–490.

Ocheja, P., Flanagan, B., Ogata, H., 2018. Connecting decentralized learning records: a blockchain based learning analytics platform. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge. ACM, Sydney, New South Wales, Australia, pp. 265–269.

Oh, I., Rho, S., Moon, S., Son, S., Lee, H., Chung, J., 2021. Creating pro-level AI for a real-time fighting game using deep reinforcement learning. IEEE Trans. Games (Early Access).

Oyelade, O.N., Ezugwu, A.E.-S., Mohamed, T.I., Abualigah, L., 2022. Ebola optimization search algorithm: A new nature-inspired metaheuristic optimization algorithm. IEEE Access 10, 16150–16177.

Park, S., Cha, H.-S., Im, C.-H., 2019. Development of an online home appliance control system using augmented reality and an SSVEP-based brain–computer interface. IEEE Access 7, 163604–163614.

Park, S.-M., Kim, Y.-G., 2022. A metaverse: Taxonomy, components, applications, and open challenges. IEEE Access 10, 4209–4251.

Pham, D.-M., 2018. Human identification using neural network-based classification of periodic behaviors in virtual reality. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces. VR, IEEE, Tuebingen/Reutlingen, Germany, pp. 657–658.

Pham, N.-Q., Kruszewski, G., Boleda, G., 2016. Convolutional neural network language models. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. ACL, Austin, Texas, pp. 1153–1162.

Pham, Q.-V., Nguyen, N.T., Huynh-The, T., Le, L.B., Lee, K., Hwang, W.-J., 2021. Intelligent radio signal processing: A survey. IEEE Access 9, 83818–83850.

Pham, Q.-V., Zeng, M., Huynh-The, T., Han, Z., Hwang, W.-J., 2022. Aerial access networks for federated learning: Applications and challenges. IEEE Netw. 36 (3), 159–166.

Qi, H., Fuin, N., Cruz, G., Pan, J., Kuestner, T., Bustin, A., Botnar, R.M., Prieto, C., 2021. Non-rigid respiratory motion estimation of whole-heart coronary MR images using unsupervised deep learning. IEEE Trans. Med. Imaging 40 (1), 444–454.

Qian, X., Chen, H., Jiang, H., Green, J., Cheng, H., Huang, M.-C., 2020. Wearable computing with distributed deep learning hierarchy: A study of fall detection. IEEE Sens. J. 20 (16), 9408–9416.

Radoff, J., 2021. The metaverse value-chain. URL https://medium.com/building-the-metaverse/the-metaverse-value-chain-afcf9e09e3a7.

Ramanathan, M., Yau, W.-Y., Teoh, E.K., 2014. Human action recognition with video data: research and evaluation challenges. IEEE Trans. Hum.-Mach. Syst. 44 (5), 650–663.

Ramu, S.P., Boopalan, P., Pham, Q.-V., Maddikunta, P.K.R., Huynh-The, T., Alazab, M., Nguyen, T.T., Gadekallu, T.R., 2022. Federated learning enabled digital twins for smart cities: Concepts, recent advances, and future directions. Sustainable Cities Soc. 79, 103663.

Rathore, M.M., Shah, S.A., 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.

Riedl, E., 2021. A look into E-commerce and more in the metaverse. URL https://www.mastercard.com/news/perspectives/2021/ecommerce-metaverse-augmented-mixed-reality/.

Rogez, G., Weinzaepfel, P., Schmid, C., 2020. LCR-net++: Multi-person 2D and 3D pose detection in natural images. IEEE Trans. Pattern Anal. Mach. Intell. 42 (5), 1146–1161.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, Munich, Germany, pp. 234–241.

Saalfeld, P., Böttcher, C., Klink, F., Preim, B., 2021. VR system for the restoration of broken cultural artifacts on the example of a funerary monument. In: Proceedings of the IEEE Virtual Reality and 3D User Interfaces. VR, IEEE, Lisboa, Portugal, pp. 739–748.

Sakhavi, S., Guan, C., Yan, S., 2018. Learning temporal information for brain-computer interface using convolutional neural networks. IEEE Trans. Neural Netw. Learn. Syst. 29 (11), 5619–5629.

Santamaría-Vázquez, E., Martínez-Cagigal, V., Vaquerizo-Villar, F., Hornero, R., 2020. EEG-Inception: A novel deep convolutional neural network for assistive ERP-based brain-computer interfaces. IEEE Trans. Neural Syst. Rehabil. Eng. 28 (12), 2773–2782.

Sharma, R., Morwal, S., Agarwal, B., Chandra, R., Khan, M.S., 2020. A deep neural network-based model for named entity recognition for Hindi language. Neural Comput. Appl. 32 (20), 16191–16203.

She, C., Dong, R., Gu, Z., Hou, Z., Li, Y., Hardjawana, W., Yang, C., Song, L., Vucetic, B., 2020. Deep learning for ultra-reliable and low-latency communications in 6G networks. IEEE Netw. 34 (5), 219–225.

She, C., Sun, C., Gu, Z., Li, Y., Yang, C., Poor, H.V., Vucetic, B., 2021. A tutorial on ultrareliable and low-latency communications in 6G: Integrating domain knowledge into deep learning. Proc. IEEE 109 (3), 204–246.

Sheikhpour, R., Sarram, M.A., Gharaghani, S., Chahooki, M.A.Z., 2017. A survey on semi-supervised feature selection methods. Pattern Recognit. 64, 141–158.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (4), 640–651.

Sun, W., Lei, S., Wang, L., Liu, Z., Zhang, Y., 2021. Adaptive federated learning and digital twin for industrial internet of things. IEEE Trans. Ind. Inf. 17 (8), 5605–5614.

Synnaeve, G., Bessière, P., 2016. Multiscale Bayesian modeling for RTS games: An application to StarCraft AI. IEEE Trans. Comput. Intell. AI Games 8 (4), 338–350.

Syrris, V., Ferri, S., Ehrlich, D., Pesaresi, M., 2015. Image enhancement and feature extraction based on low-resolution satellite data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (5), 1986–1995.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, Boston, MA, pp. 1–9.

Tang, Y., Wang, J., Wang, X., Gao, B., Dellandréa, E., Gaizauskas, R., Chen, L., 2018. Visual and semantic knowledge transfer for large scale semi-supervised object detection. IEEE Trans. Pattern Anal. Mach. Intell. 40 (12), 3045–3058.

Tanwar, S., Bhatia, Q., Patel, P., Kumari, A., Singh, P.K., Hong, W.-C., 2019. Machine learning adoption in blockchain-based smart applications: The challenges, and a way forward. IEEE Access 8, 474–488.

Tao, F., Zhang, H., Liu, A., Nee, A.Y.C., 2019. Digital twin in industry: State-of-the-art. IEEE Trans. Ind. Inf. 15 (4), 2405–2415.

Torner, J., Skouras, S., Molinuevo, J.L., Gispert, J.D., Alpiste, F., 2019. Multipurpose virtual reality environment for biomedical and health applications. IEEE Trans. Neural Syst. Rehabil. Eng. 27 (8), 1511–1520.

Tu, N.A., Huynh-The, T., Khan, K.U., Lee, Y.-K., 2019. ML-HDP: A hierarchical bayesian nonparametric model for recognizing human actions in video. IEEE Trans. Circuits Syst. Video Technol. 29 (3), 800–814.

Tunze, G.B., Huynh-The, T., Lee, J.-M., Kim, D.-S., 2020. Sparsely connected CNN for efficient automatic modulation recognition. IEEE Trans. Veh. Technol. 69 (12), 15557–15568.

Wang, S.-A., 2019. Research on fuzzy image reconstruction method based on real-time fusion technology of VR and AR. In: Proceedings of the International Conference on Virtual Reality and Intelligent Systems. ICVRIS, IEEE, Jishou, China, pp. 47–50.

Wang, X., Gao, L., Song, J., Shen, H., 2017. Beyond frame-level CNN: saliency-aware 3-D CNN with LSTM for video action recognition. IEEE Signal Process. Lett. 24 (4), 510–514.

Wang, J., Hu, Y., 2020. An improved enhancement algorithm based on CNN applicable for weak contrast images. IEEE Access 8, 8459–8476.

Wang, Q., Jiao, W., Wang, P., Zhang, Y., 2021a. Digital twin for human-robot interactive welding and welder behavior analysis. IEEE/CAA J. Autom. Sin. 8 (2), 334–343.

Wang, S., Sun, S., Wang, X., Ning, Z., Rodrigues, J.J.P.C., 2021b. Secure crowdsensing in 5G internet of vehicles: When deep reinforcement learning meets blockchain. IEEE Consum. Electron. Mag. 10 (5), 72–81.

Wang, A., Wang, W., Liu, J., Gu, N., 2019a. AIPNet: Image-to-image single image dehazing with atmospheric illumination prior. IEEE Trans. Image Process. 28 (1), 381–393.

Wang, J., Yu, L.-C., Lai, K.R., Zhang, X., 2019b. Tree-structured regional CNN-LSTM model for dimensional sentiment analysis. IEEE/ACM Trans. Audio Speech Language Process. 28, 581–591.

Wei, Y., Liang, X., Chen, Y., Shen, X., Cheng, M.-M., Feng, J., Zhao, Y., Yan, S., 2017. STC: A simple to complex framework for weakly-supervised semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (11), 2314–2320.

Weng, J., Weng, J., Zhang, J., Li, M., Zhang, Y., Luo, W., 2021. DeepChain: Auditable and privacy-preserving deep learning with blockchain-based incentive. IEEE Trans. Dependable Secure Comput. 18 (5), 2438–2455.

Wu, P., Ding, W., You, Z., An, P., 2019. Virtual reality video quality assessment based on 3D convolutional neural networks. In: Proceedings of the IEEE International Conference on Image Processing. ICIP, IEEE, Taipei, Taiwan, pp. 3187–3191.

Xu, X., Shen, B., Ding, S., Srivastava, G., Bilal, M., Khosravi, M.R., Menon, V.G., Jan, M.A., Wang, M., 2022. Service offloading with deep Q-network for digital twinning-empowered internet of vehicles in edge computing. IEEE Trans. Ind. Inf. 18 (2), 1414–1423.

Xue, Z.Y., Xiahou, K.S., Li, M.S., Ji, T.Y., Wu, Q.H., 2020. Diagnosis of multiple open-circuit switch faults based on long short-term memory network for DFIG-based wind turbine systems. IEEE J. Emerg. Sel. Top. Power Electron. 8 (3), 2600–2610.

Yang, J., Liu, T., Jiang, B., Song, H., Lu, W., 2018. 3D panoramic virtual reality video quality assessment based on 3D convolutional neural networks. IEEE Access 6, 38669–38682.

Yang, J., Wright, J., Huang, T.S., Ma, Y., 2010. Image super-resolution via sparse representation. IEEE Trans. Image Process. 19 (11), 2861–2873.

Yang, J., Xiao, L., Zhao, Y.-Q., Chan, J.C.-W., 2020. Hybrid local and nonlocal 3-D attentive CNN for hyperspectral image super-resolution. IEEE Geosci. Remote Sens. Lett. 18 (7), 1274–1278.

Yannakakis, G.N., Togelius, J., 2015. A panorama of artificial and computational intelligence in games. IEEE Trans. Comput. Intell. AI Games 7 (4), 317–335.

Yeh, C.-H., Huang, C.-H., Kang, L.-W., 2019. Multi-scale deep residual learning-based single image haze removal via image decomposition. IEEE Trans. Image Process. 29, 3153–3167.

Young, T., Hazarika, D., Poria, S., Cambria, E., 2018. Recent trends in deep learning based natural language processing [Review Article]. IEEE Comput. Intell. Mag. 13 (3), 55–75.

Zhang, W., Dong, L., Pan, X., Zou, P., Qin, L., Xu, W., 2019. A survey of restoration and enhancement for underwater images. IEEE Access 7, 182259–182279.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y., 2021. Residual dense network for image restoration. IEEE Trans. Pattern Anal. Mach. Intell. 43 (7), 2480–2495.

Zhao, L., Xu, J., Gong, C., Yang, J., Zuo, W., Gao, X., 2021. Learning to acquire the quality of human pose estimation. IEEE Trans. Circuits Syst. Video Technol. 31 (4), 1555–1568.

Zheng, C., Wu, W., Yang, T., Zhu, S., Chen, C., Liu, R., Shen, J., Kehtarnavaz, N., Shah, M., 2020. Deep learning-based human pose estimation: A survey. arXiv preprint arXiv:2012.13392.

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2020. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans. Med. Imaging 39 (6), 1856–1867.

Zhu, F., Lv, Y., Chen, Y., Wang, X., Xiong, G., Wang, F.-Y., 2020a. Parallel transportation systems: Toward IoT-enabled smart urban traffic control and management. IEEE Trans. Intell. Transp. Syst. 21 (10), 4063–4071.

Zhu, M., Sun, Z., Zhang, Z., Shi, Q., Chen, T., Liu, H., Lee, C., 2020b. Sensory-glove-based human machine interface for augmented reality (AR) applications. In: Proceedings of the IEEE 33rd International Conference on Micro Electro Mechanical Systems. MEMS, IEEE, Vancouver, BC, Canada, pp. 16–19.

Zhu, Y., Zhao, C., Guo, H., Wang, J., Zhao, X., Lu, H., 2019. Attention couplenet: Fully convolutional attention coupling network for object detection. IEEE Trans. Image Process. 28 (1), 113–126.