AI and Blockchain Empowered Metaverse for Web 3.0: Vision, Architecture, and Future Directions

Xu Zhang, Geyong Min, Tong Li, Zhan Ma, Xun Cao, and Shaowei Wang

***Abstract*—As one of the most prominent parts of the Internet, World Wide Web (WWW) has achieved great success and penetrated every area of our lives. However, the current WWW still suffers from inefficiency, growing concerns about user privacy and data ownership, and poor Quality of Experience. Therefore, we propose a promising architecture for the next-generation WWW, Web 3.0, which is underpinned by Artificial Intelligence (AI) and Blockchain-empowered Metaverse (AIB-Metaverse). Fueled by native AI, Web 3.0 based on AIB-Metaverse can provide users with a personalized experience and enable smart decision-making with high efficiency. Backed by Blockchain, the AIB-Metaverse-based Web 3.0 is decentralized, which can help users regain full control of data and protect the ownership of generated data while preserving privacy. Furthermore, this new architecture can provide a ubiquitous immersive experience to users during real-time interaction with digital avatars in the Metaverse. To verify the effectiveness of AI in Web 3.0, we propose an AI-based approach for Metaverse video delivery, which can significantly enhance the quality of immersive experience perceived by users in Web 3.0. In addition, we pinpoint the challenges faced by the proposed AIB-Metaverse-based Web 3.0 and highlight pertinent research directions in the future.**

***Index Terms*—Web 3.0, Metaverse, Edge Computing, Artificial Intelligence, Blockchain.**

I. Introduction

As one of the most prominent parts of the Internet, World Wide Web (WWW) has achieved great success and penetrated every area of our lives. Nevertheless, the current WWW has become highly centralized [1], where a massive amount of user data has been gathered, stored, and owned by a handful of tech giants. As a consequence, the data generated by users does not belong to the users themselves, and privacy leaks occur frequently. Furthermore, the current Web was not designed with native intelligence, making it inefficient to obtain the needed information from massive user-generated content (UGC) flooded in the Web. Besides, the Web cannot provide ubiquitous immersive experiences to users where their attention can be easily distracted, thus degrading user engagement.

To overcome the disadvantages of the current Web, "read-write-own" based Web 3.0 was proposed in recent years as the next generation of WWW. It is envisioned to be a decentralized web without Intermediaries [2], thus helping users regain complete control of their own data. In addition, Web 3.0 is expected to be intelligent with the capability of understanding the insights behind data, and be immersive to significantly enhance the Quality-of-Experience (QoE) perceived by users. However, the research on Web 3.0 is still in its infancy. There exists neither a commonly agreed architecture of Web 3.0 nor its implementation.

To fill the gap, we propose an architecture of Web 3.0 underpinned by Artificial Intelligence (AI) and Blockchain-empowered Metaverse (AIB-Metaverse), which is expected to equip the web with the envisioned characteristics of ubiquitous immersiveness, intelligence, and decentralization. The architecture involves two parallel worlds (*Physical World* and *Virtual World*) that inter-operate with each other, while the *Virtual World* is comprised of four layers and one core engine that work harmonically to support diversified applications. In detail, *Interaction Layer* is responsible for communicating with users and the Physical World; *Space Rendering Engine* is in charge of constructing the high-fidelity 3D environment and rendering the dynamic 3D scene with input from the *Interactive Layer*; *Smart Decision Layer* is the "brain" for the AIB-Metaverse-based Web 3.0, endowing the web with the ability to learn from data and experience by leveraging advanced AI technologies; *Secure Storage Layer* is fueled by Blockchain, which is responsible for the secure storage and trusted transactions of data in a decentralized manner whilst protecting the data ownership without a trusted third party; and *Application Layer* contains a variety of prospective applications, ranging from online education, healthcare, smart cities, to smart manufacturing. The main contributions of this article can be summarized as follows:

- We propose a promising architecture of Web 3.0, which is underpinned by AIB-Metaverse. To the best of our knowledge, this is the first architecture of AIB-Metaverse-based Web 3.0, which integrates Blockchain, AI, Digital Twin, Virtual Reality (VR)/Augmented Reality (AR)/Mixed Reality (MR), Cloud/Edge computing, and networking technologies seamlessly to help users regain full control of data, enable smart decision-making, and provide users ubiquitous immersive experience.
- We propose a Deep Reinforcement Learning (DRL) based approach to enhance the immersive experiences for users, demonstrating the effectiveness of AI in Web 3.0.
- We pinpoint the challenges faced by the proposed AIB-Metaverse-based Web 3.0, and highlight pertinent future research directions.

X. Zhang and G. Min are with the University of Exeter; T. Li is with Renmin University of China; Z. Ma, X. Cao and S. Wang are with Nanjing University. Corresponding author: X. Cao.

This work was supported by the EU Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie Grant 898588, in part by the EU Horizon 2020 INITIATE Project under Grant 101008297, and in part by NSFC Grant 62025108, 61972222, 92067206, 62202473. This article reflects only the authors' view. The European Union Commission is not responsible for any use that may be made of the information it contains.

II. Background and Limitations

A. The Evolution of the World Wide Web

As the initial stage of WWW, Web 1.0 is referred to as Syntactic Web, featured with “Read-only” web experience. During this stage, users can read the content on the websites but have little interaction with the content producers and no contribution to the content creation. In other words, the websites were static, where the information seldom flows from the users to the content producers. Representatives include personal sites and e-commerce websites, such as shopping cart applications, to showcase the information of the owner.

As the second and current stage of WWW evolution, Web 2.0 is referred to as Participative Social Web, featured with “Read-Write” web experience. During this stage, users can be the content consumer and content creator simultaneously, leading to the exponential growth of UGC. Besides, distinguished from Web 1.0 which aims at connecting users with information, Web 2.0 is highlighted with connecting users with social networking. It allows for dynamic content responsive to user input, spurring the emergence and proliferation of numerous applications (APPs) for participative communications. Representatives include the social networking platform Facebook and the video-sharing sites YouTube.

B. Limitations of the current Web 2.0

By harnessing the power of collective wisdom from billions of users via enabling real-time interaction and collaboration among users, Web 2.0 has achieved great success in all walks of life. However, it still suffers from the following limitations.

User privacy and data ownership concerns with centralized web: Although the underlying Internet is distributed, Web 2.0 is rendered to be highly centralized to a handful of extremely popular websites and applications (e.g., Facebook and Google) from tech giants. What’s worse, users on Web 2.0 cannot make transactions with each other directly. Instead, trusted intermediaries are needed, especially when the users do not trust each other. As a consequence, a massive amount of user data (e.g., age, location, or even credit card details) has been gathered, stored, and owned by these tech giants or intermediaries. It results in huge concerns regarding data ownership and data privacy, as the data may be sold like products, or even manipulated for unauthorized and illegal activities.

Inefficiency due to the lack of intelligence: During the era of Web 2.0, any user can post her/his comments or upload self-generated content to the web, which produces vast amounts of data. As a sequence, all kinds of information from different people with diversified thoughts flood the whole network. The information may even contain malicious comments and rumors (e.g., injection of bleach can protect you against the COVID-19 virus [3]), which makes it tougher for users to obtain useful accurate information effectively. Actually, if the web has native intelligence to analyse the massive data and optimize the process of the Web (*e.g.* access to the web via the dynamic and bandwidth-limited Internet), the efficiency of the Web can be significantly improved.

Poor QoE due to the lack of ubiquitous immersive experience: During the era of Web 2.0, users mainly access the web via browsers or mobile applications in the format of 2 dimensional (2D) content. It achieves great success due to the low cost and short time to create 2D content, and the low requirements on the viewing devices regarding the computing capability. However, it cannot provide immersive experience to users, where users’ attention can be easily distracted thus degrading user engagement. Although the current VR/AR/MR technologies can provide a sort of immersive experience to users, they still suffer from numerous obstacles that should be overcome in order to provide ubiquitous access to users with heterogeneous viewing devices and dynamic network conditions.

III. AIB-Metaverse-based Web 3.0

A. Upcoming Web 3.0

To cope with the challenges faced by the current Web 2.0, “Read-Write-Own” based Web 3.0 is envisioned as the next generation of WWW to eliminate the threat of censorship, helping users regain complete control of their own data. It is expected to be backed by Blockchain technologies to achieve decentralization without intermediaries, thus no central point of control exists and no single tech giant can control the identities of others.

In addition, by harnessing the power of semantics and AI, it is expected that the machines in Web 3.0 are equipped with the capability to understand the meaning behind data. This means that the Web can find what users are actually interested in, improving the efficiency of web searching. Furthermore, ubiquitous immersive experience is anticipated by advancing VR/AR/MR technologies.

B. Architecture of Web 3.0 underpinned by AIB-Metaverse

Inspired by the promising Metaverse [4], we propose an architecture of AIB-Metaverse-based Web 3.0, which involves two parallel worlds (Physical World and Virtual World) that inter-operate with each other. On one hand, the status of the Physical World is captured by diversified sensors, and uploaded to the Virtual Space to construct and update the digital avatars and environment. On the other hand, the Virtual Space accepts the input from the users and the data from the Physical World, renders the scene in a real-time manner, and gives feedback to the Physical Space, which turns into actions that the objects in the Physical World should execute.

As shown in Fig. 1, the Virtual Space should encompass four layers and one core engine to unleash the full potential of Web 3.0. In detail, when a user accesses a Web 3.0 application available in the *Application Layer*, the *Space Rendering Engine* will first obtain the needed data from the *Secure Storage Layer* and reconstruct the 3D scene based on historical interaction of the user. During the interaction, the commands from the user, such as the head movements and eye-tracking from the VR/AR/MR helmet, the finger tapping from the smartphone, or the mouse clicks from the laptop, are captured by the interactive terminal and sent to the *Interactive Layer*, which interacts with the user via its digital avatar. The

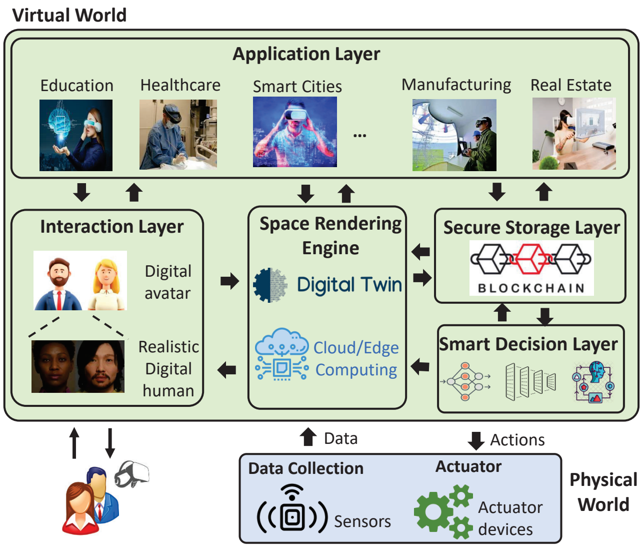

Fig. 1: The architecture of Metaverse-empowered Web 3.0

commands are further input to the *Space Rendering Engine*, which takes advantage of the AI models in the *Smart Decision Layer* to update the scene accordingly. Based on the updated scene, the *Interactive Layer* will generate and compress the video, and stream the video to the user. Note that the data generated by the users, such as the interaction logs, can be stored in the *Secure Storage Layer* for future exploitation, if necessary.

This architecture equips the web with the following characteristics, which are exactly what is envisioned for Web 3.0.

1) **Decentralization:** The web is backed by Blockchain technology to avoid the data being controlled by a handful of tech giants. Known for its transparency and immutability, Blockchain can play a paramount role in protecting digital assets in Web 3.0.
2) **Intelligence:** AI is envisioned to be "brain" of the web to provide personalized user experience and enable smart decision-making. It helps to understand the input from users, create automated 3D avatars, and recommend the needed information to enable tailored experiences for different individuals.
3) **Ubiquitous immersiveness:** In the AIB-Metaverse-based Web 3.0, users have 3D avatars as digital representatives of themselves in the virtual world. As such, a sense of trust can be created for users by providing ubiquitous immersive experiences during the interaction with other avatars or the environment.

C. Components

This subsection introduces the key components of the proposed AIB-Metaverse-based Web 3.0.

1) Interaction Layer: This layer is responsible for communicating with the users and the Physical World. On one hand, it interacts with users via their 3D digital avatars with the objective to provide ubiquitous immersive experiences to them. To enable users to have smooth access to the immersive Web anytime anywhere, no matter whether the computing powers on their devices are strong or not, the computation-intensive tasks (*e.g.,* the scene rendering) can be offloaded from users' devices to the cloud/edge, while the rendered 3D scene is streamed to users' devices in the form of videos. Note that HTTP-based video delivery is adopted to reuse the existing network infrastructure, thus reducing the upfront cost.

On the other hand, the Virtual World can inter-operate with the Physical World via the *Interactive Layer*. Specifically, the real-time status of physical infrastructure is collected by sensors, and then uploaded to the Physical World with the *Interactive Layer*. If applicable, the status will be analysed by the *Smart Decision Layer*, such as status prediction and anomaly detection. Then corresponding actions will be fed back to the Physical World and executed by actuator devices.

2) Space Rendering Engine: This engine is in charge of constructing the high-fidelity 3D environment and rendering the dynamic 3D scene with the input from the *Interactive Layer*. Distinguished from Web 2.0, data in Web 3.0 is visualized in 3D format by the engine, which offers a faster and more effective way to communicate insights. Specifically, *Space Rendering Engine* leverages Digital Twin to clone objects from the Physical World into their digital counterparts in the Virtual World, which can act as the building blocks to further construct the high-fidelity 3D environment. In addition, driven by the real-time status collected by the *Interactive Layer*, the digital counterparts can also be utilized to simulate the objects' behaviors, thus facilitating the status forecast and configuration optimization of the objects in the Physical World. These digital counterparts will be stored in the *Secure Storage Layer* to ensure the data integrity and protect the ownership.

As space rendering is computation-intensive, the *Space Rendering Engine* will exploit the scalable and powerful computing capability of cloud computing to reduce operational complexity. In addition, to enhance the user experience, edge computing will be introduced to place computational rendering tasks to network edge (e.g., base stations and roadside units) in close proximity, thus satisfying the stringent requirements of delay-sensitive Web 3.0 services.

3) Smart Decision Layer: Powered by advanced AI, this layer plays a vital role of "brain" for the AIB-Metaverse-empowered Web 3.0, endowing the web with the ability to learn from data and experience. Without doubts, huge amount of data will be generated by users and services in Web 3.0, which provides the *Smart Decision Layer* the potential to train powerful AI models for personalized user experience, smart decision making, and web process optimization. For example, to achieve automated digital avatars, which can smartly respond to the input from a user, the *Smart Decision Layer* leverages multimodal deep learning to understand the user by analyzing speech and visual modalities. In detail, the *Smart Decision Layer* will maintain three types of AI models to support various tasks in Web 3.0, as specified below.

- The first type is supervised learning models, which deal with labeled data and determine the output for an input based on historical input-output pairs. For example, the *Smart Decision Layer* can pre-train mature classification networks (*e.g.* GoogLeNet) for object recognition, which could be applied as the backbone network to train a specific model to recognize a physical object when constructing its digital counterpart via Digital Twin technologies.

- The second type is unsupervised learning models, which deal with unlabeled data and learn the hidden patterns without human intervention. For instance, the *Smart Decision Layer* can leverage the k-prototypes algorithm for user behavior pattern analysis, which could enhance the personalized user experience in Web 3.0.
- The last type is reinforcement learning models, which learn to achieve a cumulative reward through a sequence of decisions in a complex environment by striking a balance between exploration and exploitation. For example, the *Smart Decision Layer* can utilize A3C [5] to train advanced DRL models for adaptive Metaverse video streaming, maximizing users' experience within the dynamic network.

4) Secure Storage Layer: Backed by Blockchain, this layer is responsible for the secure storage and trusted transactions of data in a decentralized manner whilst protecting the data ownership without a trusted third party. Distinguished from Web 2.0 where data are collected and stored by the tech giants, data generated by users in Web 3.0 are stored in Blockchain, which is managed by a peer-to-peer network maintained by users themselves. Furthermore, the peers will synchronize their local ledgers to reach an agreement regarding the state of current data with consensus protocols, such as proof-of-stake (PoS), which make it impossible to alter the data in the Blockchain. In detail, the *Secure Storage Layer* will adopt the following measures to protect data ownership and privacy while guaranteeing the trustiness of transactions.

- To realize the trusted transactions of data among anonymous users without a trusted intermediary, Smart Contacts are adopted by the *Secure Storage Layer*. Supported by Smart Contacts, the state-changing operations from any participant in Web 3.0 will be verified by the participant in the Blockchain network, thus guaranteeing the trustiness of transactions.
- To facilitate the ownership and authenticity of users' data, the *Secure Storage Layer* allocates non-fungible tokens (NFTs) to assets atop Blockchain, such as the trained AI models in *Smart Decision Layer*. When there is a dispute over the ownership of an asset, its NFT is leveraged to determine the origin and the current owner in seconds [6].
- To protect the users' privacy, the *Secure Storage Layer* combines off-chain storage and on-chain personal data access control to realize trusted storage and secure sharing of personal data on the Blockchain. Specifically, ciphertext policy attribute-based encryption (CP-ABE) is leveraged to allow that users who conform to the set of attributes can access the data.

5) Application Layer: With the Virtual World, the AIB-Metaverse-based Web 3.0 becomes a perfect place to simulate scenarios that are expensive, dangerous, harmful, or even infeasible in the Physical World. It can also provide immersive and personalized experiences to users, thus can accommodate various prospective applications, ranging from online education, healthcare, smart cities, manufacturing, to real estate. An eye-catching example is online education [7], where interactive learning environments can be tailored for students with different backgrounds and in different learning institutions. Specifically, an automated Metahuman can act as a teacher to train students with professional knowledge and skills. It can also play the role of a virtual student to attend the discussion with students. Apart from routine teaching activities, counterproductive situations can also be reenacted in the Virtual World to educate the students when managing a problematic behavior in a school [8].

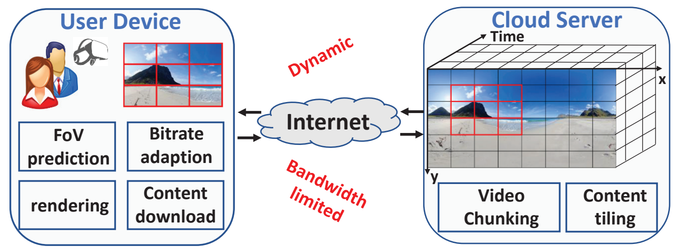

Fig. 2: The developed prototype system.

D. Case Study

In this section, we demonstrate the vital role of AI in Web 3.0, by verifying its effectiveness in enhancing the efficiency of the Metaverse video delivery and improving the immersive experience for users in Web 3.0. Specifically, we set up a prototype system to support Field of View (FoV)-based Metaverse video delivery, which only transmits the part of the video within a user's FoV rather than the whole omnidirectional 360° Metaverse video, thus reducing the bandwidth waste within the bandwidth-limited Internet [9].

As shown in Fig. 2, a user wears a VR helmet to view Metaverse videos hosted by a remote server. To enable adaptive bitrate within the dynamic Internet, the server would first split the whole video into small chunks (with a duration of 1s), divide each chunk into $n_l \times n_w$ tiles, and encode each tile into m different representations with different bitrates. In this way, the user can predict its FoV to determine which tiles are required, select the bitrate of each tile with a designed DRL model, download the specific tiles from the server, and render the downloaded tiles for real-time viewing, thus achieving high quality of experience within the dynamic Internet.

Distinguished from the existing DRL models for 360° video streaming, which ignore the possible error of FoV prediction, our DRL model takes the viewing probability of a tile into consideration. As shown in Fig. 3, the observed state includes the past k tile throughputs $(x_t, x_{t-1}, ..., x_{t-k+1})$, the rebuffering time for the past k tiles $(R_t, R_{t-1}, ..., R_{t-k+1})$, the candidate sizes of the next tile $(s_1, s_2, ..., s_m)$, the bitrate of the last tile b_t, the number n_t of remaining tiles within the same chunk, the number c_t of remaining chunks for the Metaverse video, and the viewing probability of the tile p_k. Regarding the reward, a higher average bitrate r_t^{avg} of a chunk is preferred, while the quality variation r_t^{intra} among all the tiles within the FoV, the quality variation r_t^{inter} between the neighboring chunks, and the rebuffering time τ_t for each chunk

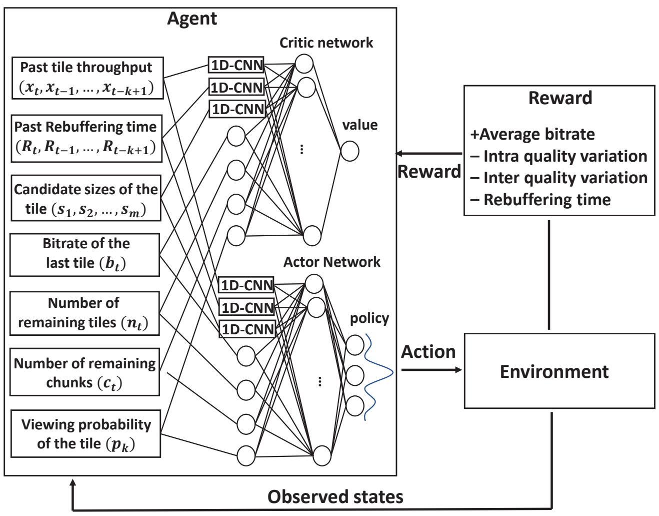

Fig. 3: The proposed DRL model.

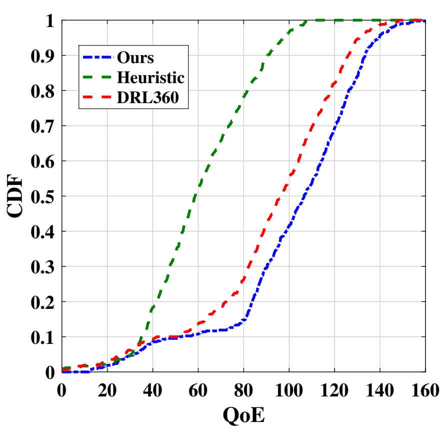

Fig. 4: Performance comparison with existing methods.

are given penalties. In other words, the reward function is defined as

$$r_t = r_t^{avg} - \mu_1 r_t^{intra} - \mu_2 r_t^{inter} - \mu_3 \tau_t. \quad (1)$$

In this way, the proposed model will try to maximize the video quality, enhance the video quality smoothness, and minimize the rebuffering time perceived by users. Trained with the A3C, the AI model will determine which one from the m different representations should be selected.

In our experiment, each chunk is divided into 8×4 tiles while each tile is encoded into 6 representations, that is, n_l equals 8, n_w equals 4, and m equals 6. Besides, the user device utilizes a Hidden Markov Model [10] to predict the FoV. And the past 8 statuses are input to a 1D-CNN with 128 filters, and other states are input to a fully connected layer with 128 neurons, that is k equals 8. As users are more sensitive to the video rebuffering than the video quality smoothness, we set μ_1 to 0.5, μ_2 to 0.5, and μ_3 to 5. Similarly to [9], we utilize open-access network trace datasets to train our AI models, including a 4G-LTE network dataset, an HSDPA dataset, and a broadband dataset from FCC.

We compare our proposed AI models with an existing heuristic method (referred as to *Heuristic*) [11] and the state-of-the-art DRL method for 360° video streaming (referred as to *DRL360*) [12]. The performance metric is defined as the QoE perceived by users, which is also the reward function utilized to train our AI models. As shown in Fig. 4, our proposed approach outperforms the two benchmark methods in terms of the QoE perceived by users. It is worth noting that both the two AI-based approaches can significantly improve the QoE, which demonstrates the effectiveness of AI in enhancing the immersive experiences of users on Web 3.0. This is mainly because AI models can directly learn useful insights for the bitrate decision-making from historical data without any assumption about the environment. In addition, our approach advances the state-of-the-art DRL model for 360° video streaming by taking into consideration the inaccuracy of FoV prediction.

IV. Challenges and Future Directions

Currently, the AIB-Metaverse-based Web 3.0 is in its nascent stage with many challenges to be overcome.

A. *Ubiquitous Interaction with ultra-low latency requirements*

Users are expected to interact with the AIB-Metaverse-based Web 3.0 via heterogeneous viewing devices, such as VR/AR/MR headsets, smartphones, etc. On one hand, some devices are only equipped with limited computing, storage, and networking resources, and are incapable to execute computation-extensive rendering tasks to view high-quality metaverse videos. On the other hand, ultra-low motion-to-photon latency should be satisfied to prevent cybersickness when having VR/AR/MR headsets [13]. In other words, the scene should be rendered, streamed, and displayed on users' headsets within tens of milliseconds after the head movements.

To tackle the above challenges, smart task offloading mechanisms should be designed to migrate some tasks from viewing devices to edge servers or cloud centres. In addition, considering the spatial and temporal coherence of the users' FoV, along with the locational and behavioral coherence among different users [14], we can take advantage of proactive edge caching to reduce the latency for the Metaverse video transmission.

B. *Ultra-high-fidelity Metahuman rendering to cross the Uncanny Valley*

Metahuman, or realistic digital human, is promising to act as the digital representative of a user to enhance the immersive experience. With powerful cloud/edge computing, a user can create his/her own Metahuman by defining the facial features, the body shape, the haircut style, and so on. However, human-like visual representations that imperfectly resemble actual human beings will confuse the brain's visual processing systems, making our sentiment plummet deeply into negative emotional territory, which is known as Uncanny Valley.

To leap over the uncanny valley, many service providers just abandon the option of Metahuman. They remove the human-like features from the avatar by using a mechanical appearance and motion or directly designing cartoon-like faces [15]. However, it can degrade the user-perceived experience by creating a sense of unreality. We believe that with the development of AI technologies, it is expected to generate ultra-high-fidelity Metahuman in the near future.

C. Explainable AI for trustworthy decision-making

As the "brain" of the Metaverse-empowered Web 3.0, the *Smart Decision Layer* leverages AI for decision-making and service recommendation. However, many AI models are kind of black boxes whose inputs and operations are not visible or understandable to Metaverse operators. This black-box nature often leads to non-transparent decisions that lack interpretability and reliability. It makes users hardly believe that the AI-fueled web can be trustworthy and will cause no harm. For instance, the *Space Rendering Engine* would render the virtual space after recognizing the input from users. However, if the AI approaches utilized for the space rendering are not trusted, the rendered virtual space will also be doubted.

Therefore, we should design trustworthy AI to render trusted Web 3.0 services. It should align with the ethical, legal and social values of the society. In addition, We should ensure that the trained AI model is explainable, robust, fair, and with high precision. For example, the trained AI model should be robust enough to cope with emergency situations which are not included in the training data set. In case an AI model does not work in an extreme situation, we will have contingency plans to guarantee the smooth running of Web 3.0 services.

D. Scalable and resource-efficient Blockchain for frequent data sharing

To achieve decentralized web, data generated by users is stored in Blockchain to protect the data ownership and authenticity without a trusted third party. However, the current Blockchain technology is incompetent for Web 3.0 as it suffers from scalability and resource-efficiency issues due to low response speed and high resource demands. For instance, with the Proof-of-Work consensus protocol, Bitcoin, the leading cryptocurrency atop Blockchain, can only process 3-7 transactions per second while each transaction should wait for about 10 minutes to reach the global confirmation. Furthermore, nodes should keep a full record of data to ensure the data is immutable, that is, 324GB in the storage space. Nevertheless, Web 3.0 should have the capability to support millions of users to share data simultaneously. What's worse, most of the users are using resource-limited devices, such as smartphones or AR helmets, which can hardly store the full record of data with the ever-growing size.

Therefore, revolutionary Blockchain consensus should be designed in the future to speed up transaction processing whilst keeping the Web 3.0 services reliable. To make Blockchain resource-efficient, Blockchain sharding can be a promising research direction by splitting the processing of transactions among smaller groups of nodes. With Blockchain sharding, the communication, computation, and storage overhead of processing transactions can be significantly reduced.

V. Conclusion

We propose the very first architecture of AIB-Metaverse driven Web 3.0, which integrates Blockchain, AI, Digital Twin, AR/VR/MR, Cloud/Edge computing, and networking technologies seamlessly to help users regain full control of data, enable smart decision-making, and provide users ubiquitous immersive experience. To verify the effectiveness of AI, we propose a DRL-based approach and demonstrate that it can enhance the performance of Metaverse video delivery. In addition, we pinpoint the challenges faced by the proposed Metaverse-based Web 3.0 and highlight pertinent research directions in the future.

References

[1] R. W. Gehl, "Distributed centralization: Web 2.0 as a portal into users' lives," *Lateral*, no. ateral 1, 2012.

[2] F. A. Alabdulwahhab, "Web 3.0: the decentralized web blockchain networks and protocol innovation," in *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1–4, IEEE, 2018.

[3] van Der Linden *et al.*, "Inoculating against fake news about covid-19," *Frontiers in psychology*, vol. 11, p. 566790, 2020.

[4] R. Cheng *et al.*, "Will Metaverse be NextG Internet? Vision, Hype, and Reality," *arXiv preprint arXiv:2201.12894*, 2022.

[5] V. Mnih *et al.*, "Asynchronous methods for deep reinforcement learning," in *International conference on machine learning*, pp. 1928–1937, PMLR, 2016.

[6] Q. Wang *et al.*, "Non-fungible token (NFT): Overview, evaluation, opportunities and challenges," *arXiv preprint arXiv:2105.07447*, 2021.

[7] H. Duan *et al.*, "Metaverse for Social Good: A University Campus Prototype," in *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 153–161, 2021.

[8] S. Mystakidis, "Metaverse," *Encyclopedia*, vol. 2, no. 1, pp. 486–497, 2022.

[9] Z. Jiang *et al.*, "Reinforcement learning based rate adaptation for 360-degree video streaming," *IEEE Transactions on Broadcasting*, vol. 67, no. 2, pp. 409–423, 2021.

[10] S. Xie *et al.*, "Perceptually optimized quality adaptation of viewport-dependent omnidirectional video streaming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 146–160, 2020.

[11] S. Petrangeli *et al.*, "An HTTP/2-based adaptive streaming framework for 360 virtual reality videos," in *The 25th ACM international conference on Multimedia*, pp. 306–314, 2017.

[12] Y. Zhang *et al.*, "DRL360: 360-degree Video Streaming with Deep Reinforcement Learning," in *IEEE Conference on Computer Communications*, pp. 1252–1260, 2019.

[13] E. Bastug *et al.*, "Toward interconnected virtual reality: Opportunities, challenges, and enablers," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 110–117, 2017.

[14] M. Zink *et al.*, "Scalable 360 video stream delivery: Challenges, solutions, and opportunities," *Proceedings of the IEEE*, vol. 107, no. 4, pp. 639–650, 2019.

[15] L. Kugler, "Crossing the uncanny valley," *Communications of the ACM*, vol. 65, no. 8, pp. 14–15, 2022.

Xu Zhang (x.zhang6@exeter.ac.uk) is a Marie Sklodowska-Curie research fellow at the University of Exeter, U.K. His research interests include artificial intelligence, cloud/edge computing, and multimedia networks.

Geyong Min (g.min@exeter.ac.uk) is a professor at the University of Exeter, U.K. His research interests include computer networks, wireless communications, parallel and distributed computing, and performance engineering.

Tong Li (tong.li@ruc.edu.cn) is an associate professor at Renmin University of China (RUC) and a research fellow at the Metaverse Research Center of RUC. His research interests include networking, distributed systems and big data.

Zhan Ma (mazhan@nju.edu.cn) is a professor at Nanjing University, China. His research interests focuses on the neural video coding and smart cameras.

Xun Cao (caoxun@nju.edu.cn) is a professor at Nanjing University, China. His research interests include computational photography, image-based modeling and rendering, and VR/AR systems.

Shaowei Wang (wangsw@nju.edu.cn) is a professor at Nanjing University, China. His research interests include communications and networking, operations research and machine learning.