



Roadmap toward the metaverse: An AI perspective

Shenghui Cheng,¹ Yue Zhang,¹ Xiaofei Li,¹ Lin Yang,¹ Xin Yuan,^{1,*} and Stan Z. Li^{1,*}

¹AI Division, School of Engineering, Westlake University, Hangzhou 310024, China

*Correspondence: xyuan@westlake.edu.cn (X.Y.); stan.zq.li@westlake.edu.cn (S.Z.L.)

Received: June 9, 2022; Accepted: July 24, 2022; Published Online: August 1, 2022; <https://doi.org/10.1016/j.xinn.2022.100293>

© 2022 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Citation: Cheng S., Zhang Y., Li X., et al., (2022). Roadmap toward the metaverse: An AI perspective. *The Innovation* 3(5), 100293.

Truth must be told, but not much—a hybrid society of real and virtual is coming. Metaverse,¹ rise recently, is attracting significant attention from academia to industry. A metaverse is a network of three-dimensional (3D) virtual worlds focused on social connection. Bearing the outbreak of the coronavirus 2019 pandemic, people are physically isolated, which triggered the growth of the metaverse. Different from existing work, this commentary targets the roadmap of the metaverse from an artificial intelligence (AI) perspective. First, we blueprint a roadmap for this digital cyberspace transformation, including immersion creation, hardware support, text interpretation, audio processing, connection construction, economy operation, and security protection. Second, at each phase of the roadmap, we address the status as well as the advanced technologies to provide in-depth views accordingly. The whole pipeline is illustrated from an AI² perspective, as shown in Figure 1, and we believe that AI is playing an increasingly important role in core techniques toward this technological singularity.

Immersion creation refers to the technique that provides human beings with immersive feelings by constructing a 3D virtual world. Inspired by computer vision, graphics, and visualization techniques, immersion creation includes generating virtual scenes in the metaverse and displaying them to end users. AI has revolutionized these techniques recently. For instance, the scene generation process has been significantly speeded up to nearly real time. However, the scenes generated by these computer-vision-based methodologies are limited by the pre-

defined elements used in AI algorithms, which would preclude the immersive feeling of human interaction with a digital human in the metaverse since these elements are different from the real ones, but we desire both parties to share the same scene and thus the same feeling of the environment. This brings challenges in sensing, sampling, and scene generation in the metaverse. One solution is to capture real scenes and use those images or videos to generate new ones and update the scenes in the metaverse with a short period of time, possibly in real time eventually. To this end, computational imaging has provided a promising solution to capture scenes efficiently in a low-cost, low-bandwidth manner,³ along with AI, computational imaging may develop rapidly in metaverse-related applications in the future to provide a higher sense of immersion.

Hardware denotes the end-user devices, such as the brain-computer interface, robotics, and virtual reality or augmented reality headsets as well as glass-free 3D devices, which determine the quality of the user's immersive experience. These devices have been revolutionized many times in history,⁴ taking the AR headset as an example, it was designed from macro- to micro-optics and now is heading toward nano-optics, and the size has been reduced dramatically. However, one main challenge is the design of large-scale diffractive devices used in glasses, and AI is now helping with and speeding up the design in a number of ways, from accelerating the iterations to proposing new solutions. The AI accelerator, such as neural-inspired AI chips, designed to accelerate AI and its applications,

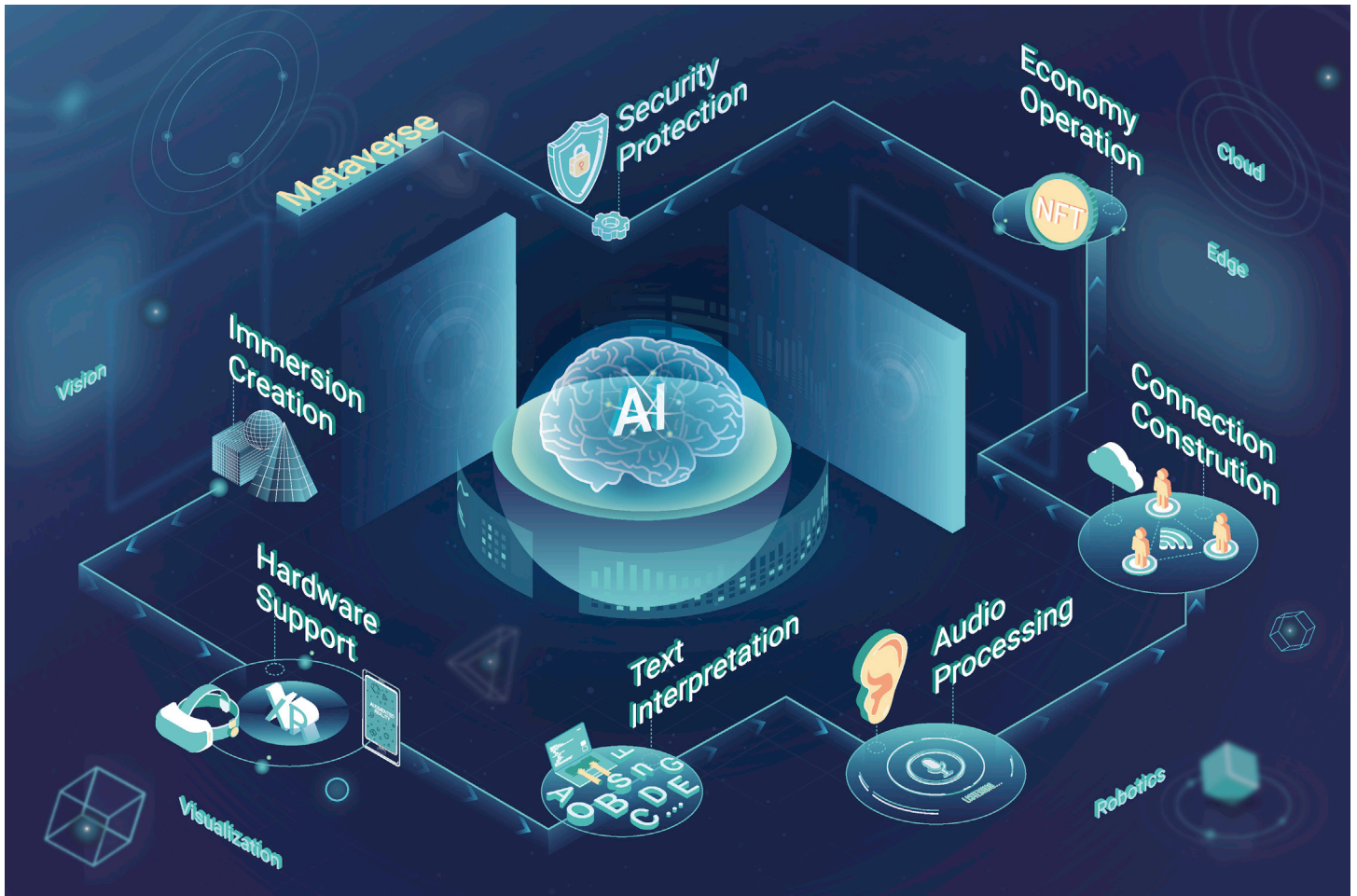


Figure 1. Metaverse roadmap supported by AI The key steps are marked with icons surrounded by AI. Some essential techniques are floating around

is growing tremendously now and will evolve to task-specific chips to be used in the metaverse. We believe this is a trend in optical or other devices' design and is not limited to the metaverse.

Text interpretation regards to text generation and text for communication purpose. The metaverse brings convenience for recording activities in business and social life, which potentially facilitates the role of natural language processing,⁵ a research domain widely supported by AI. Most natural language processing applications in the metaverse focus on personal assistance and business meeting analysis: from the personal side, new dialogue systems could benefit book-keeping of social activities, enabling a personal assistant to give clear instructions, and from a business perspective, it keeps track of meeting minutes, where AI algorithms generate summaries and answer questions concerning specific details. However, in a hybrid society, how to communicate with people of different backgrounds even living in different eras could be an interesting AI-powered application.

Audio processing aims for the rendering of auditory immersion. Voice is considered one important interface for humans to enter the metaverse, and it is also the main interaction model between entities (avatars, digital humans, or even non-human objects) in the metaverse. Voice processing consists of two major tasks: automatic speech recognition and text to speech (or speech synthesis), which transforms voice signals to text, or vice versa. Together with the language-understanding techniques, automatic speech recognition and text to speech enable entities in the metaverse to understand the message and intention of others and to speak as if they were in the real world. In addition, audio signals could be rendered as binaural signals, from which humans can sense the location of the sound source and the presence of an enclosed space; in other words, to have the sense of auditory immersion. For auditory immersion in the metaverse, many challenges need to be further overcome, such as separating metaverse sound from real-world sound, customized speech synthesis, complex 3D soundscape generation, etc., which highly rely on advanced signal-processing and machine-learning techniques. Sound is ubiquitous in the metaverse, and we believe that AI is playing a leading role in audio management in the metaverse.

Connection construction in the metaverse includes network connection and social construction to get connected with others. Network connections require high-speed network to transfer massive amounts of data, accounting for the obstacles in latency, bandwidth, and consistency. AI overcomes the obstacles in various aspects including traffic planning, routing and classification, congestion control, quality of service, and quality of experience management. After being networked, the society emerges but faces challenges like behavior communication or community management. AI is capable of optimizing social group management by matching people with their desired communities so that the whole society will be divided spiritually rather than physically. As the core purpose of the connection, social activity is a requisite of some recognition techniques, such as face recognition for meeting people and gesture recognition for behavior interaction, as well as pose tracking, texture restoration, and blur correction—all of which are common cases where AI excels.

Economy operation is for exchanging virtual goods through a blockchain digital system. With the exponential growth of virtual economy, some challenges have arisen, like large-scale asset management and fraudulent-transaction detection. AI empowers the virtual economy by generating and managing digital assets, monitoring transactions among huge databases, and more. New classifiers and mechanisms supported by AI verify the authenticity of transactions via a decentralized blockchain infrastructure to improve security posture. AI-generated non-fungible tokens and AI-assisted asset tagging increase the efficiency of asset generation and management. In the future, AI can build a more reliable technology-based virtual economy system.

Security protection corresponding to the security framework and privacy protection becomes extremely challenging when tremendous virtual devices are connected. This connection is not that reliable and gives space to get attacked, e.g., identity theft and spying, which are common problems that security faces. To be protected, the security system has to require frequent authentications when accessing the service related to virtual devices, which results in time-consuming and other problems. New mechanisms are demanding for authentication with alternative modalities, such as biometric authentication driven by muscle movements or eye gazes, as well as seamless authentication. None of these would happen without AI. Moreover, countless activities and interactions are recorded, creating a crisis of personal privacy. AI is capable of privacy protection through algorithms that automatically and dynamically detect user privacy preferences from diverse contexts in the metaverse. An AI-powered security system will emerge in the metaverse.

Besides, other techniques like robotics, cloud computing, the Internet of Things, and decision-making are also important components for the metaverse. Likewise, AI also contributes significantly to them—for example, with the help of reinforcement learning, decision-making demonstrates great potential since it enables AI to make decisions through huge amounts of data with various factors and has now become a new trend.

AI not only advances the above technologies but is itself evolving rapidly. Current components such as immersion creation and hardware techniques are not really "intelligent" by themselves but are composed "sensing/sampling + AI algorithms" to claim to be intelligent. Looking forward, the next step of AI in the metaverse should move from "make the hardware intelligent by AI algorithms" to "build intelligent hardware by AI algorithms," where the latter is a hybrid mode and focuses on end-to-end solutions. In this solution, other important techniques in AI such as reinforcement learning, federated learning, and few-shot learning will probably lead to new revolutions.

A new era driven by AI is emerging. The digital human produced by AI will increase dramatically, and a digital second life is appearing—a society combined with real and digital humans is launching. It is likely that other species or organisms will be reborn and that a "creator-verse" will turn up. Immortal technology—living forever—will eventually become true. At the current stage, it is more from the inorganic side where everything is generated with digital computers; however, the society will continue, and people will eventually reach the third-verse or even the multi-verse, where the organic world and the inorganic world may merge. AI will speed up this revolution.

REFERENCES

- Cheng, S.H. (2022). *Metaverse: Concepts, Technologies and Ecology* (China Machine Press).
- Xu, Y.J., Liu, X., Cao, X., et al. (2021). Artificial intelligence: a powerful paradigm for scientific research. *Innovation* **2**, 100179.
- Altmann, Y., McLaughlin, S., Padgett, M.J., Goyal, V.K., Hero, A.O., and Faccio, D. (2018). Quantum-inspired computational imaging. *Science* **361**, 6403. <https://doi.org/10.1126/science.aat2298>.
- Lee, L.K., Braud, T., Zhou, P.Y., et al. (2021). All one needs to know about metaverse: a complete survey on technological singularity, virtual ecosystem, and research agenda. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2110.05352>.
- Zhang, Y., and Teng, Z.Y. (2021). *Natural Language Processing: A Machine Learning Perspective* (Cambridge University Press).

ACKNOWLEDGMENTS

We would like to thank the support from the Research Center for Industries of the Future (RCIF) at Westlake University, Westlake Foundation (2021B1501-2) and the funding from Lochn Optics.

DECLARATION OF INTERESTS

The authors declare no competing interests.