Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Perspective:

Sora for foundation robots with parallel intelligence: three world models, three robotic systems[*]

Lili FAN[†1], Chao GUO[2], Yonglin TIAN[2], Hui ZHANG[3], Jun ZHANG[4], Fei-Yue WANG[†‡2]
[1]*School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*
[2]*State Key Laboratory for Management and Control of Complex Systems, Institute of Automation,*
Chinese Academy of Sciences, Beijing 100190, China
[3]*School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China*
[4]*School of Electrical Engineering, Wuhan University, Wuhan 430072, China*
[†]E-mail: lilifan@bit.edu.cn; feiyue.wang@ia.ac.cn

Received Mar. 3, 2024; Revision accepted Apr. 6, 2024; Crosschecked Apr. 16, 2024; Published online Apr. 22, 2024

https://doi.org/10.1631/FITEE.2400144

This paper outlines the initial steps and basic framework for developing foundation/infrastructure robots/robotics based on foundation models and parallel intelligence, as well as the potential applications of new artificial intelligence (AI) techniques such as AlphaGO, ChatGPT, and Sora.

1 Introduction

Following the emergence of foundation models such as ChatGPT (Wang JC et al., 2023; Zhou et al., 2023) and Sora, there is a focus on developing technology that leverages these advancements. Sora is a new generative AI system that can create videos up to 60 s in length based on textual instructions. These videos are highly realistic, containing detailed scenery, vivid character expressions, and complex camera movements. Sora not only understands the elements of prompts given by users, but also comprehends how these elements interact in the physical world. Thanks to its high-fidelity results and the cost savings compared to human work, Sora (or related models in the future) will likely influence content creation in media and entertainment industries, such as film and video gaming. Furthermore, there is potential to revolutionize education and training, for instance, with the quick generation of visual tutorials. Additionally, because of their effectiveness in translating textual (or voice) commands into visual effects, such methods could bring about significant changes in embodied AI applications and autonomous driving technologies. Several initial reports have been written on Sora (Li X et al., 2024; Qin et al., 2024; Wang FY et al., 2024; Yu et al., 2024); however, the potential impact of foundation models on robotics has not been explored in depth.

In the context of AI development, there is a desire to orient novel technologies towards greater societal benefit (Wang FY, 2021; Ye et al., 2022; Fan LL et al., 2023a). Recent technologies such as foundation models are being used to transform scientific and industrial research into more virtual or data-driven forms. This can be thought of as augmenting (or in some cases replacing) traditional methods with artificial counterparts, for instance, leveraging machine

---

[‡] Corresponding author

[*] Project supported by the National Key Research and Development Program of China (No. 2023YFB3209802) and the China Postdoctoral Science Foundation (No. 2023M740264)

ORCID: Lili FAN, https://orcid.org/0000-0002-8767-7123; Fei-Yue WANG, https://orcid.org/0000-0001-9185-3989

learning to make discoveries in natural sciences, using artificial manufacturing to research new materials or designs, or introducing digital beings and robots to parallel biological beings (Ye and Wang, 2022; Wang FY, 2023b). Such developments have the potential to reveal new ideas with greater ease, reduce the risk involved in human testing of novel technologies, and create cost savings compared to manual approaches. Therefore, there is potential for a paradigm shift in science and technology, where humans work in parallel on scientific research with increasingly advanced machine intelligence such as ChatGPT or Sora (Wang FY and Wang, 2024), or similarly on the engineering of new technology. With this in mind, we discuss the potential impact of newly developed foundation models on robotics and automation, and propose a framework by which robotic systems could be automatically trained by such models.

2 Big robotic world models: three worlds, three robots

The philosopher Karl POPPER famously proposed a "Three Worlds" model of our reality, which consists of physical, mental, and artificial worlds. In a similar manner, we define three types of robots, physical, human, and digital, as a means to describe the integration of AI technologies (Fig. 1). Accordingly, we develop the concept of "foundation robot" based on foundation models and parallel intelligence (Miao et al., 2023), which could potentially enhance automation of tasks and development of machine intelligence.

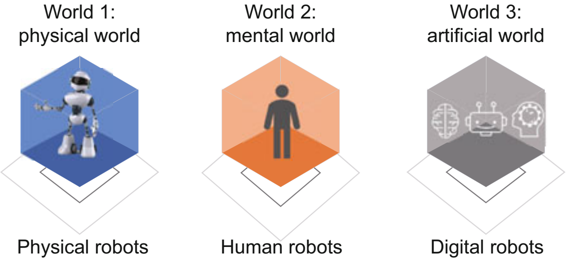

Fig. 1 Three worlds, three types of robots

2.1 Physical robots

A physical robot possesses a tangible physical form, consists of mechanical structures and actuators, uses electronic devices such as sensors, and has the abilities to move and interact with its environment. Such robots perform various actions in the real world, with many familiar examples from domains such as manufacturing, healthcare, and transportation. These actions can range from simple repetitive tasks to highly complex operations. For the design and manufacturing of such robots, factors in the physical environment such as mechanics, dynamics, and friction need to be considered. In increasingly complex settings, physical robots require perceptual and adaptive capabilities to function successfully; AI provides a means to enhance these capabilities. Physical robots that use modern foundation models and parallel intelligence exhibit stronger learning capabilities, more efficient task execution, greater adaptability across different domains, and higher performance and accuracy, thus bringing numerous advantages to a range of applications (Wang XX et al., 2023).

2.2 Human robots

In his 1651 book *Leviathan*, Thomas HOBBES wrote, "Life is but a motion of limbs. For what is the heart, but a spring; and the nerves, but so many strings; and the joints, but so many wheels, giving motion to the whole body, such as was intended by the Artificer?" (Hobbes, 2017). In a modern context, this statement can be thought of as a description of the similarities between humans and robots. In fact, many of these similarities are quite intuitive. First, human behavior and thinking are influenced by factors such as biology (genetics) and sociology (environment, culture, etc.), akin to how robots are affected by their underlying computer code and their external environment. Second, human behavioral responses in specific situations can be based on prior experiences or innate biological mechanisms, resembling the predictable and interpretable responses of robots. Third, some human behaviors and habits can become automated, similar to how robots execute preset programs. For instance, routine tasks like driving or washing the dishes can essentially be performed without conscious thought by humans. Additionally, note that both humans and robots possess the capability for technological augmentation; for example, humans enhance their senses and thinking with technology (e.g., using glasses to enhance vision and using computers to enhance cognitive abilities). This is similar to how robots use sensors and actuators, or improved computer code, to augment their functionalities.

2.3 Digital robots

Digital robots, also known as software robots, automation software, or AI agents, are virtual entities that execute specific tasks through computer programs. These robots do not exist physically, but simulate physical or human behavior to perform tasks in an automated fashion. Digital robots are commonly employed to handle work with high volume, repetition, and data input requirements, and integration across multiple legacy systems. Such robots can range from simple programming scripts to systems that use AI to automatically learn and optimize their performance. In business and industrial settings, digital robots are used for tasks such as office automation, customer service, and data analysis. They are also often used in robotic process automation (RPA), offering benefits such as reduction of human errors, improvement of time value, rapid execution of repetitive tasks, and continuous operation. As digital robots become increasingly advanced, they could be considered a kind of digital human (Malik and Brem, 2021).

3 Foundation models for foundation robots via parallel robotics

Based on a framework of parallel robotics and foundation models, human robots, digital robots, and physical robots can cooperate to achieve process and knowledge automation, thereby optimizing solutions to various real-world tasks (Guo et al., 2022, 2023a). We define this cooperative system as a “foundation robot,” and explore its future utility.

The concept of “parallel robotics” was proposed by Prof. Fei-Yue WANG in 2015, integrating the parallel system theory (Wang FY, 2004), the artificial-computational-parallel (ACP) approach (Wang FY, 2007), and general principles of robotics. Its goal is to address the high experimental costs and low learning efficiency of traditional robotic systems (Bai et al., 2017). This methodology establishes virtual artificial systems for robotics, where computational experiments are conducted to expand the exploration space and improve learning efficiency, thereby enhancing the adaptability and efficacy of robotic systems. Additionally, it integrates humans into the training of robotic systems, where they provide demonstrations, supervision, and evaluation. Parallel robotics extends robots from the cyber-physical space to the cyber-physical-social space, thus enabling iterative optimization (or collaborative learning) between virtual and physical robotic systems. Leveraging this methodology in combination with foundation models, in what we call a “foundation robot,” could lead to further advancements.

A potential operational pipeline for foundation robots is illustrated in Fig. 2. Human robots (or actual humans) propose overall requirements to digital robots and are responsible for training, supervising, and evaluating the digital and physical robots. This could involve human demonstration, value alignment based on human feedback, plan approval, and other quality checks. Digital robots, which are constructed based on specialized foundation models, engage in task planning and scheduling, and delegate tasks to other robots according to the generated plans. If the desired application is physical, physical robots would respond and execute tasks in space until the goal is achieved. Digital and physical robots could operate autonomously for most of the time (say 80%) while providing necessary monitoring information to human robots. In a minority of the time (say 15%),

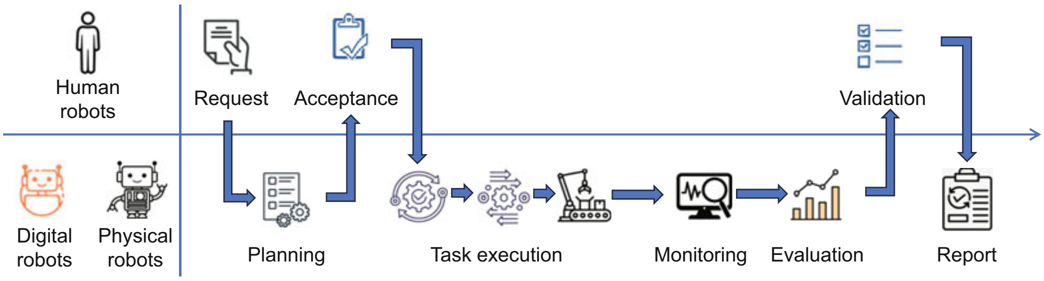

Fig. 2 Framework for foundation robots based on parallel robotics and foundation models

the system could operate in a parallel mode, with human robots intervening to guide the other robots to resolve exceptions. In exceptional cases (say 5%), an expert/emergency mode could be triggered, with humans directly handling issues until the system returns to a normal autonomous operation (Wang FY, 2022). Therefore, through collaboration between the robotic members, foundation robots may achieve process automation, as well as accumulation and evolution of knowledge.

Machine learning models play a vital role in the operation of a foundation robotic system. Following leaps in algorithmic intelligence made by models like AlphaGO (or its successor MuZero), and leaps in linguistic intelligence made by foundation models like ChatGPT, Sora has demonstrated a development in imaginative intelligence (Guo et al., 2023b). Trained on massive multi-modal data and general tasks, foundation models can learn generalized representations of the surrounding world and causal relationships (Fan LL et al., 2023b, 2024b, 2024c). They could therefore lay a foundation of downstream learning tasks for digital and physical robots. Foundation models can be applied to generate various scenarios from multiple perspectives and predict future trends, thus enabling large-scale computational experiments and testing in artificial systems. Furthermore, foundation models serve as excellent channels for human–machine interaction, providing the status of systems to humans in a friendly manner, and facilitating communication between various robot types. Foundation models like Sora could empower foundation robots to evolve into a kind of "infrastructure robot," which would form the core infrastructure of future intelligent robotic systems, providing a framework to build solutions for complex problems.

4 Digital robotic schools with curriculum learning for parallel intelligence development

To improve the learning and generalization abilities of robotic systems, we propose the concept of "digital robotic school." Here, diverse scenarios and tasks are constructed, and curriculum learning approaches (Wang X et al., 2022) are adopted to gradually instruct robots. A main motivation for this framework is to develop robots with parallel intelligence, which is defined as a combination of descriptive, predictive, and prescriptive intelligence (Miao et al., 2023). A proposed framework for a digital robotic school is shown in Fig. 3. This school can be thought of as three parts: digital teaching, curriculum learning, and parallel learning.

In digital teaching, large language models (LLMs) such as ChatGPT (Wang FY et al., 2023) and artificial intelligence generated content (AIGC) models like Sora could be used for the design and generation of scenarios and curricula, thereby providing a foundation to train and evaluate robots (Li X et al., 2023; Tian et al., 2023a). The so-called "digital teachers" could be based on foundation models and generative intelligent technologies, to operate alongside human teachers. Digital teaching could be implemented to cultivate various types of digital robots, and serve as a foundation for future human–machine interactions. This would potentially be relevant for applications in the metaverse, where realistic

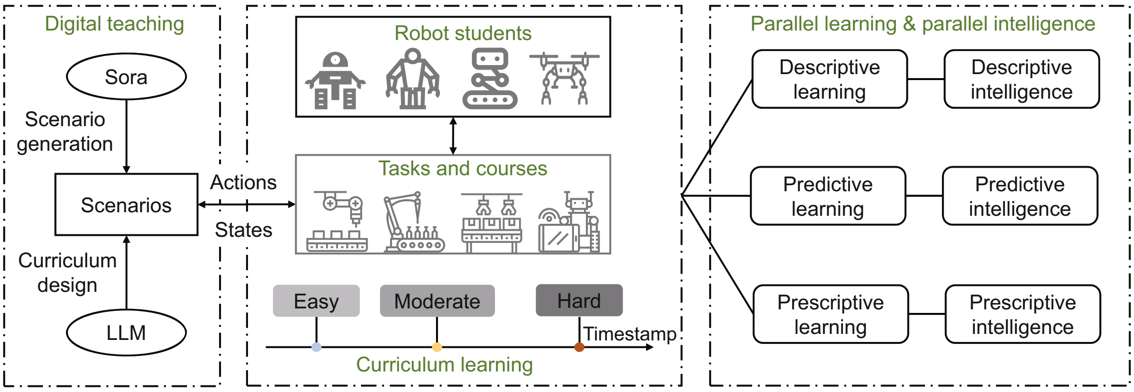

Fig. 3 Proposed framework for a digital robotic school (LLM: large language model)

integration of humans into virtual environments is of paramount importance. Furthermore, digital teaching would come with great potential for real-time visualization and instant availability of knowledge. Currently or in the near future, ChatGPT (or similar LLMs) could likely facilitate the design of a comprehensive training environment akin to a digital robotic school, thanks to its ability to understand the intricacies of specific tasks and prompts. This design process might involve the identification of tasks for learning, the skills necessary to accomplish such tasks, and the generation of scenarios that incrementally teach such skills, effectively forming a scaffolding structure for instruction. In the context of visual learning, Sora (or related AIGC models) would enable the creation of vivid, interactive scenarios that mimic real-world complexities. Such scenarios would provide robots with learning more similar to what a human experiences, potentially leading to the development of adaptable, intelligent robotic behaviors that can handle multifaceted challenges. The task understanding of ChatGPT and the scenario generation of Sora could also be leveraged in tandem within digital robotic schools, providing a robust training foundation to impart robots with nuanced and versatile abilities, and pave the way for parallel intelligence (Yang et al., 2023).

By organizing flexible federated scenarios and courses (Wang FY et al., 2021; Tian et al., 2022), curriculum learning can be implemented to train robots for various tasks. In the context of a digital robotic school, robots could undergo a systematic and structured process of skill acquisition, which closely mirrors human educational strategies. This could begin with foundational tasks to establish basic capabilities, and progress towards more complex challenges as a robot's proficiency increases. A step-wise approach facilitates gradual development of skills for various tasks such as perception (Tian et al., 2023b) and planning (Shen et al., 2024), and also allows for dynamic adjustment of the curriculum based on the robot's performance, thus promoting comprehensive and efficient learning.

To bridge the gap between simulated environments and real-world applications, parallel learning (Li L et al., 2017; Miao et al., 2023; Fan LL et al., 2024b) could be adopted within digital robotic schools. This methodology integrates descriptive, predictive, and prescriptive learning into a cohesive framework, establishing parallel systems that enhance the robot's learning efficiency and generalization ability. Essentially, parallel learning leverages the mutual dependencies between data, knowledge, and actions in both real and virtual spaces. Accordingly, robots undergo a series of self-boosting and self-adaptive stages, mirroring the complexity of real-world scenarios in a controlled, simulated environment. This methodology not only expedites the learning process, but also improves the robot's ability to adapt to new situations, thus bolstering its descriptive and predictive intelligence. Furthermore, with parallel execution based on hybrid (virtual–real) interactions, feedback can be obtained from the hybrid space and then implemented to optimize real applications. This can make robots more capable of bridging the gap between simulation-based training and real-world execution, thus fulfilling the desired goal of prescriptive intelligence.

5 Conclusions: towards infrastructure robots

The development of foundation robots for specific functions across various fields could lead to more generalized robotic systems. Considering future improvement of computing power and further optimization of algorithmic models, foundation robots will likely possess enhanced learning and reasoning abilities, enabling them to extract more accurate insights from larger-scale data. Consequently, these increasingly advanced foundation robots may form basic infrastructures for a given field, upon which more specific applications could be built in the future; we refer to this concept as an "infrastructure robot" or "infrastructure robotic system." For example, in the context of transportation, we may see that a foundation robot working for a specific vehicle (like a car) becomes generalized to work for another transportation infrastructure (Wang FY, 2023a; Wang FY and Lv, 2023). This infrastructure robot could then be used as a basis for further transportation applications. Accordingly, we would likely see infrastructure robots for educational, medical, logistical, and a multitude of service and industrial applications. With increasing advancements in AI technologies, foundation and infrastructure robots could become the new basic machines upon which smart cities and societies are developed.

Contributors

Lili FAN drafted the paper. Chao GUO, Yonglin TIAN, and Fei-Yue WANG helped organize the paper. Hui ZHANG, Jun ZHANG, and Fei-Yue WANG revised and finalized the paper.

Conflict of interest

Fei-Yue WANG is an executive associate editor-in-chief of *Frontiers of Information Technology & Electronic Engineering*, and he was not involved with the peer review process of this paper. All the authors declare that they have no conflict of interest.

References

Bai TX, Wang S, Shen Z, et al., 2017. Parallel robotics and parallel unmanned systems: framework, structure, process, platform and applications. *Acta Autom Sin*, 43(2):161-175 (in Chinese). https://doi.org/10.16383/j.aas.2017.y000002

Fan LL, Cao DP, Zeng CX, et al., 2023a. Cognitive-based crack detection for road maintenance: an integrated system in cyber-physical-social systems. *IEEE Trans Syst Man Cybern Syst*, 53(6):3485-3500. https://doi.org/10.1109/TSMC.2022.3227209

Fan LL, Zeng CX, Li YJ, et al., 2023b. GRC-Net: Fusing GAT-based 4D Radar and Camera for 3D Object Detection. SAE Technical Paper 2023-01-7088.

Fan LL, Zeng CX, Meng ZL, et al., 2024a. A secured vehicle brain: DAO-based collaborative perception and decision-making systems for intelligent vehicles in CPSS. *IEEE Trans Intell Veh*, 9(1):52-54. https://doi.org/10.1109/TIV.2023.3341698

Fan LL, Wang JH, Chang YM, et al., 2024b. 4D mmWave radar for autonomous driving perception: a comprehensive survey. *IEEE Trans Intell Veh*, early access. https://doi.org/10.1109/TIV.2024.3380244

Fan LL, Zeng CX, Wang YT, et al., 2024c. Social radars: finding targets in cyberspace for cybersecurity. *IEEE/CAA J Autom Sin*, 11(2):279-282. https://doi.org/10.1109/JAS.2024.124251

Guo C, Bai TX, Wang X, et al., 2022. ShadowPainter: active learning enabled robotic painting through visual measurement and reproduction of the artistic creation process. *J Intell Robot Syst*, 105(3):61. https://doi.org/10.1007/S10846-022-01616-1

Guo C, Dou Y, Bai TX, et al., 2023a. ArtVerse: a paradigm for parallel human–machine collaborative painting creation in metaverses. *IEEE Trans Syst Man Cybern Syst*, 53(4):2200-2208. https://doi.org/10.1109/TSMC.2022.3230406

Guo C, Lu Y, Dou Y, et al., 2023b. Can ChatGPT boost artistic creation: the need of imaginative intelligence for parallel art. *IEEE/CAA J Autom Sin*, 10(4):835-838. https://doi.org/10.1109/JAS.2023.123555

Hobbes T, 2017. Leviathan. Penguin Classics. Available from https://www.penguin.co.uk/books/263235/leviathanby-hobbes-thomas/9780141395098 [Accessed on Jan. 20, 2024].

Li L, Lin YL, Zheng NN, et al., 2017. Parallel learning: a perspective and a framework. *IEEE/CAA J Autom Sin*, 4(3):389-395. https://doi.org/10.1109/JAS.2017.7510493

Li X, Tian YL, Ye PJ, et al., 2023. A novel scenarios engineering methodology for foundation models in metaverse. *IEEE Trans Syst Man Cybern Syst*, 53(4):2148-2159. https://doi.org/10.1109/TSMC.2022.3228594

Li X, Miao QH, Li LX, et al., 2024. Sora for senarios engineering of intelligent vehicles: V&V, C&C, and beyonds. *IEEE Trans Intell Veh*, early press. https://doi.org/10.1109/TIV.2024.3379989

Malik AA, Brem A, 2021. Digital twins for collaborative robots: a case study in human-robot interaction. *Robot Comput Integr Manuf*, 68:102092. https://doi.org/10.1016/j.rcim.2020.102092

Miao QH, Lv YS, Huang M, et al., 2023. Parallel learning: overview and perspective for computational learning across Syn2Real and Sim2Real. *IEEE/CAA J Autom Sin*, 10(3):603-631. https://doi.org/10.1109/JAS.2023.123375

Qin R, Wang FY, Zheng XL, et al., 2024. Sora for computational social systems: from counterfactual experiments to artificiofactual experiments with parallel intelligence. *IEEE Trans Comput Soc Syst*, 11(2):1531-1550. https://doi.org/10.1109/TCSS.2024.3373928

Shen ZP, Zhou GZ, Huang HL, et al., 2024. Convex optimization-based trajectory planning for quadrotors landing on aerial vehicle carriers. *IEEE Trans Intell Veh*, 9(1):138-150. https://doi.org/10.1109/TIV.2023.3327263

Tian YL, Wang JG, Wang YT, et al., 2022. Federated vehicular Transformers and their federations: privacy-preserving computing and cooperation for autonomous driving. *IEEE Trans Intell Veh*, 7(3):456-465. https://doi.org/10.1109/TIV.2022.3197815

Tian YL, Li X, Zhang H, et al., 2023a. VistaGPT: generative parallel transformers for vehicles with intelligent systems for transport automation. *IEEE Trans Intell Veh*, 8(9):4198-4207. https://doi.org/10.1109/TIV.2023.3307012

Tian YL, Zhang XJ, Wang X, et al., 2023b. ACF-Net: asymmetric cascade fusion for 3D detection with LiDAR point clouds and images. *IEEE Trans Intell Veh*, early access. https://doi.org/10.1109/TIV.2023.3341223

Wang FY, 2004. Artificial societies, computational experiments, and parallel systems: a discussion on computational theory of complex social-economic systems. *Compl Syst Compl Sci*, 1(4):25-35 (in Chinese). https://doi.org/10.13306/j.1672-3813.2004.04.002

Wang FY, 2007. Toward a paradigm shift in social computing: the ACP approach. *IEEE Intell Syst*, 22(5):65-67. https://doi.org/10.1109/MIS.2007.4338496

Wang FY, 2021. Parallel philosophy: origin and goal of intelligent industries and smart economics. *Bull Chin Acad Sci*, 36(3):308-318 (in Chinese). https://doi.org/10.16418/j.issn.1000-3045.20210301001

Wang FY, 2022. Parallel management: the DAO to smart ecological technology for complexity management intelligence. *Acta Autom Sin*, 48(11):2655-2669 (in Chinese). https://doi.org/10.16383/j.aas.c220773

Wang FY, 2023a. Infrastructural vehicles: the TAO from infrastructural intelligence to foundational mobility and services. *IEEE Trans Intell Veh*, 8(12):4667-4670. https://doi.org/10.1109/TIV.2023.3341692

Wang FY, 2023b. New control paradigm for Industry 5.0: from big models to foundation control and management. *IEEE/CAA J Autom Sin*, 10(8):1643-1646. https://doi.org/10.1109/JAS.2023.123768

Wang FY, Lv C, 2023. Foundation vehicles: from foundation intelligence to foundation transportation for future mobility. *IEEE Trans Intell Veh*, 8(10):4287-4291. https://doi.org/10.1109/TIV.2023.3326636

Wang FY, Wang YT, 2024. Digital scientists and parallel sciences: the origin and goal of AI for science and science for AI. *Bull Chin Acad Sci*, 39(1):27-33 (in Chinese). https://doi.org/10.16418/j.issn.1000-3045.20231212004

Wang FY, Qin R, Chen YZ, et al., 2021. Federated ecology: steps toward confederated intelligence. *IEEE Trans Comput Soc Syst*, 8(2):271-278. https://doi.org/10.1109/TCSS.2021.3063801

Wang FY, Miao QH, Li X, et al., 2023. What does ChatGPT say: the DAO from algorithmic intelligence to linguistic intelligence. *IEEE/CAA J Autom Sin*, 10(3):575-579. https://doi.org/10.1109/JAS.2023.123486

Wang FY, Miao QH, Li LX, et al., 2024. When does Sora show: the beginning of TAO to imaginative intelligence and scenarios engineering. *IEEE/CAA J Autom Sin*, 11(4):809-815. https://doi.org/10.1109/JAS.2024.124383

Wang JC, Tang Y, Hare R, et al., 2023. Parallel intelligent education with ChatGPT. *Front Inform Technol Electron Eng*, early access. https://doi.org/10.1631/FITEE.2300166

Wang X, Chen YD, Zhu WW, 2022. A survey on curriculum learning. *IEEE Trans Patt Anal Mach Intell*, 44(9):4555-4576. https://doi.org/10.1109/TPAMI.2021.3069908

Wang XX, Yang J, Wang YT, et al., 2023. Steps toward Industry 5.0: building "6S" parallel industries with cyber-physical-social intelligence. *IEEE/CAA J Autom Sin*, 10(8):1692-1703. https://doi.org/10.1109/JAS.2023.123753

Yang J, Wang X, Tian YL, et al., 2023. Parallel intelligence in CPSSs: being, becoming, and believing. *IEEE Intell Syst*, 38(6):75-80. https://doi.org/10.1109/MIS.2023.3284694

Ye PJ, Wang FY, 2022. Parallel population and parallel human—a cyber-physical social approach. *IEEE Intell Syst*, 37(5):19-27. https://doi.org/10.1109/MIS.2022.3208362

Ye PJ, Wang X, Zheng WB, et al., 2022. Parallel cognition: hybrid intelligence for human-machine interaction and management. *Front Inform Technol Electron Eng*, 23(12):1765-1779. https://doi.org/10.1631/FITEE.2100335

Yu HK, Liu XY, Tian YL, et al., 2024. Sora-based parallel vision for smart sensing of intelligent vehicles: from foundation models to foundation intelligence. *IEEE Trans Intell Veh*, early access. https://doi.org/10.1109/TIV.2024.3376575

Zhou J, Ke P, Qiu XP, et al., 2023. ChatGPT: potential, prospects, and limitations. *Front Inform Technol Electron Eng*, early access. https://doi.org/10.1631/FITEE.2300089