








## Perspective

# The Journey/DAO/TAO of Embodied Intelligence: From Large Models to Foundation Intelligence and Parallel Intelligence

By Tianyu Shen , Jinlin Sun , Shihan Kong , Yutong Wang , Member, IEEE, Juanjuan Li , Member, IEEE, Xuan Li , Fei-Yue Wang , Fellow, IEEE

**T**HE tremendous impact of large models represented by ChatGPT [1]–[3] makes it necessary to consider the practical applications of such models [4]. However, for an artificial intelligence (AI) to truly evolve, it needs to possess a physical “body” to transition from the virtual world to the real world and evolve through interaction with the real environments. In this context, “embodied intelligence” has sparked a new wave of research and technology, leading AI beyond the digital realm into a new paradigm that can actively act and perceive in a physical environment through tangible entities such as robots and automated devices [5].

The concept of embodied intelligence was first proposed by Turing in 1950 [6]. It focuses on creating intelligent entities that can autonomously learn and evolve with a combination of software and hardware based on the interaction between machines and the physical world, thus devoting to solve more real-world problems. However, in the past few decades, embodied intelligence has not made much progress because the technologies were not yet sufficient to support its development. Today, interdisciplinary technologies such as robotics, large

models and deep learning, reinforcement learning, computer vision, computer graphics, natural language understanding, cognitive science, etc., have changed this situation.

In particular, with the popularity of “large models + robots” combination, both the academic and industrial sectors have introduced tangible achievements on embodied intelligence. For example, Google unveiled SayCan of Everyday Robot by integrating robots and conversation models, enabling robots to complete a long-sequence task consisting of 16 steps with the help of a large language model [7]. UC Berkeley presents LM Nav by leveraging three large models (ViNG as visual navigation model, GPT-3 as large language model and CLIP as visual language model) for teaching robots to reach the designated destination based on language commands without map navigation [8].

To this end, we can consider some fundamental issues of embodied intelligence and explore future development directions. Compared to traditional AI systems, embodied intelligence emphasizes that robots can better adapt to complex environments by direct interaction, thus performing a variety of tasks in real physical world. Given that large models adhere to the “big problems, big models” paradigm, there lacks the ability to analyze real-time data/trends and highly specialized knowledge, thus a “small problems, big models” paradigm is advocated. In this way, are advanced foundation models in specific fields and big data for their training needed? Also, how to coordinate human beings and more robots or digital humans in real production? We will study such issues by discussing the history, kernel and prospects of embodied intelligence in the following sections.

In view of the above, this study will effectively integrate advanced foundation models and the parallel intelligence methodology, enabling deep collaboration among the robots, digital and biological human beings for real-world applications. It also serves the parallel intelligence industry through autonomous, parallel, and expert/emergency modes, achieving the “6S” goals: Safety, Security, Sustainability, Sensitivity, Service, and Smartness [9].

## The History of Embodied Intelligence

- **Embodied intelligence vs disembodied intelligence.** Intelligence accumulates via learning and evolution in the natural environment [10]. The embodied cognition and

Citation: T. Shen, J. Sun, S. Kong, Y. Wang, J. Li, X. Li, and F.-Y. Wang, “The journey/DAO/TAO of embodied intelligence: From large models to foundation intelligence and parallel intelligence,” *IEEE/CAA J. Autom. Sinica*, vol. 11, no. 6, pp. 1313–1316, Jun. 2024. (Corresponding authors: Xuan Li and Fei-Yue Wang)

T. Shen is with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: tianyu.shen@buct.edu.cn).

J. Sun is with the School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China (e-mail: jsun@ujs.edu.cn).

S. Kong is with the State Key Laboratory for Turbulence and Complex Systems, Department of Advanced Manufacturing and Robotics, College of Engineering, Peking University, Beijing 100871, China (e-mail: kongshihan@pku.edu.cn).

Y. Wang and J. Li are with State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: yutong.wang@ia.ac.cn; juanjuan.li@ia.ac.cn).

Xuan Li is with the Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: lix05@pcl.ac.cn).

F.-Y. Wang is with the State Key Laboratory for Management and Control of Complex Systems, Chinese Academy of Sciences, Beijing 100190, and also with Faculty of Innovation Engineering, Macau University of Science and Technology, Macau 999078, China (e-mail: feiyue.wang@ia.ac.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JAS.2024.124407

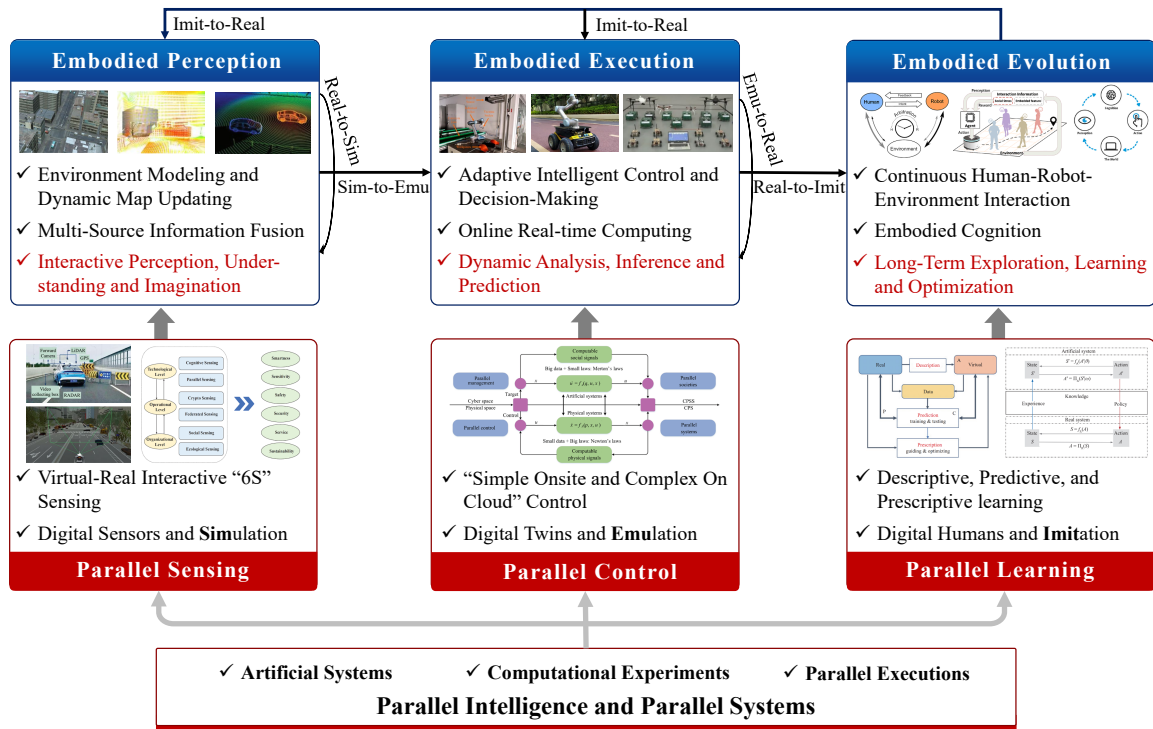


Fig. 1. The kernel of embodied intelligence and the correspondence with parallel intelligence methodology.

intelligence can be rapidly acquired by the struggles, interactions, and collaborations between agents' morphologies and their environment. However, classical artificial intelligence primarily focuses on the disembodied intelligence, constructing diverse pure digital models to further elucidating what are the input data. Some state-of-art models might even be well in understanding multimodal data, such as language, vision, and voice. Note that robots have real bodies interacting to their environment frequently. Therefore, embodied intelligence, which is superior to disembodied intelligence, is prerequisite for diverse robots in reality.

- **Large models and multimodal learning.** Human has the capability to leverage multiple modalities of perception data collectively in order to engage themselves fitting with dynamic and unconstrained environment. Inspired by the aforementioned, large models [11] and multimodal learning [12] are organized together to imitate human perception more tightly. The scale of deep neural network has increased to 1.7 trillion parameters not long ago, surpassing human in some ways probably. Meanwhile, the transferability, efficiency, robustness, and interpretability of transformer-based multimodal learning methods have been improved via enlarging models. In this climate, the time has come that large models and multimodal are utilized in the field of embodied intelligence.
- **Digital humans and robotics.** The impact of robotics and digital humans on embodied intelligence is far-reaching and multi-layered. Robotic entities make embodied intelligence more relevant to real-world needs and challenges through actual interaction with the environment, includ-

ing perception and action in physical space. In addition, the integration of perception and action in a single entity by robotics drives the synergistic optimization of perception and decision by the embodied intelligence system. This integration motivates the embodied intelligence system to understand perceptual information better and adjust decision strategies accordingly. It is meaningful to note that, in addition to robotics, digital humans can also be applied to embodied intelligence systems for interacting with the environment [13]. The construction of digital humans can help realize more intelligent and flexible embodied intelligence systems and improve their adaptability in complex environments.

### The Kernel of Embodied Intelligence

Embodied intelligence is an intelligent system that is coupled by "ontology" and "intelligent agent" and enables performing tasks in complex environments. Our founding Editor-in-Chief, Prof. Fei-Yue Wang, establishes the Parallel Intelligence and Parallel System [14], based on his study on shadow systems [15] in early 1990s, for dealing with high-uncertainty, diverse, and complex systems to be agile, focused, and convergent with the ACP method and virtual-real interaction. This paper considers that embodied intelligence consists of three kernel modules including embodied perception, embodied execution, and embodied evolution, which can be supported by relevant methods in parallel intelligence society, as shown in Fig. 1.

- **Embodied perception and parallel sensing.** Embodied perception is the most essential part for achieving embodied intelligence. In contrast to previous perception based on computer vision, embodied perception should

be characterized by dynamic, multimodal, and interactive to meet the uncertain, diverse and complex environments. Firstly, environment modeling and dynamic map updating are required with integrating some foundation models and incorporating special domain knowledge. Also, multi-source information fusion is indispensable with multi-sensor integration and multimodal learning methods [12]. Additionally, interactive perception, understanding and imagination are emphasized for suppressing multi-source uncertainty in open environments such as the uncertainty of observation conditions, uncertainty of dynamic changes and uncertainty of environmental interference. Such pipeline could be a real-to-simulation process. The parallel sensing methodology [16], [17] provides a comprehensive guidance and implementation way for the purpose of this module. Specifically, the parallel sensing defines a constitution of real physical sensors and virtual digital sensors, and allows virtual-real sensor interaction with some simulation technologies [18] to boost the adaption in various scenarios and conserve the energy, ultimately achieving the goal of “Cognitive, Parallel, Crypto, Federated, Social and Ecologic” 6S sensing.

- **Embodied execution and parallel control.** When the environment is perceived, you will definitely think about how to deal with it and emulate the process in your mind so as to achieve better execution in real tasks. Similarly, following the embodied perception, embodied execution stage with a emulation-to-real process is required. The basic demand for the embodied execution is characterized by adaptive, real-time, and dynamic for achieving the uncertain, diverse and complex tasks. Some necessary investigations include adaptive intelligent control and decision making, online real-time computing, as well as dynamic analysis, inference and prediction. It is expected that the execution can adapt to various changes in real-world events, which is very challenging. Parallel control [19], [20] paradigm with digital twins [21], [22] has provided some thoughts. With the presence of digital twin models, parallel control enables the controlled system and the control guidance to be mathematically symmetrical by introducing the time derivative of the control vector, which makes it easier to introduce adversarial game, machine learning, artificial intelligence etc., for constructing the “simple onsite, complex on cloud” control.
- **Embodied evolution and parallel learning.** The human body is the most prominent embodied intelligent agent. The way that humans learn deserves in-depth consideration for providing imitation reference for the evolution of embodied intelligence, which is called as embodied evolution. Some key points for the embodied evolution include continuous human-robot-environment interaction, embodied cognition, as well as long-term exploration, learning and optimization. The preliminary idea in this regard is inspired by parallel learning methodology [23], [24], which is proposed to diminish the obstacles for practical application of theoretical achievements. In contrast to traditional machine learning theories, parallel learning views intelligent agents and environments as a cohesive

system rather than two opposing systems, by constructing an artificial system parallel to the real system and designing three main components including descriptive learning, predictive learning, and prescriptive learning. In addition, digital humans can also be applied to embodied evolution for interacting with the environment, for realizing more intelligent and flexible embodied intelligence systems.

### *The Prospects of Embodied Intelligence*

We believe that the development of embodied intelligence connects potentially with foundation intelligence and parallel intelligence, to achieve Turing’s initial vision of embodied intelligence. It is worth noting that we insist on the goal of embodied intelligence not to replace humans with robots imitating humans, but rather to augment biological humans in the mental world with both robots in the physical world and digital humans in the artificial world to, and to form new-era parallel intelligence [25] through interacting with each other and building parallel systems in the three worlds.

- **Foundation models and foundation intelligence.** For embodied intelligence, it is basic and essential to master embodied knowledge, which can be achieved through domain-specific foundation models [11]. The progress made in foundation model technologies, as demonstrated by tools such as ChatGPT, is expected to speed up the incorporation of embodied knowledge. At the same time, the robots, digital humans, and biological humans are organized (interaction, alliance and collaboration) within the DAOs (Decentralized Autonomous Organizations and Decentralized Autonomous Operations) framework and HOOS (Human-Oriented Operating Systems) [26]. The various elements in CPSS (cyber, physical, and social spaces) interact with each other by means of foundation intelligence, such as federated intelligence, cognitive intelligence, ecological intelligence, and DeSci [27], through autonomous, parallel, and expert/emergency modes (APeM) [28], [29], achieving the “6S” goals.
- **Parallel intelligence and parallel population.** As shown in Fig. 1, it is demonstrated that how parallel intelligence and parallel systems support the different stages of embodied intelligence, which is a great test bed for new-era parallel intelligence methodology and technologies. In the future, it is essential for us to cultivate a parallel community with parallel population through the interaction and connection of three categories of individuals across three worlds. That is, biological humans comprise less than 5%, and robots comprise less than 15%, and digital humans comprise over 80% through working modes of APeM, so as to guarantee the human-oriented attribute with a projected improvement of 96000% in effectiveness and efficiency in the future [25].
- **DAOs/TAOs and ecological intelligence.** Trust and attention are both crucial for embodied intelligence to realize ecological intelligence. When attention has been greatly addressed by the foundational intelligence technologies represented by foundation models, trust still necessitates the incorporation of decentralized technologies like blockchain, smart contracts, and Decentralized

Autonomous Organizations and Operations or TRUE Autonomous Organizations and Operations [30], [31]. The integration of them will enhance the ability of intelligent agents to interact harmoniously and sustainably within diverse environmental systems. The data generation and learning processes of embodied intelligence can be meticulously recorded using blockchain technology to enhance their traceability and transparency. This also enables three categories of humans to organize themselves into DAOs/TAOs, where services and tasks provided by them can be organized and operated in a more fair and efficient manner. Furthermore, the mechanisms and strategies within DAOs/TAOs can be standardized and codified, allowing for the execution of these processes through smart contracts. This method ensures a high level of sustainability in the functioning and management of embodied intelligence.

#### ACKNOWLEDGMENT

This work was partially supported by the National Natural Science Foundation of China (62302047, 62203250) and the Science and Technology Development Fund of Macau SAR (0093/2023/RIA2, 0050/2020/A1).

#### REFERENCES

- [1] Y. Wang, X. Wang, X. Wang, J. Yang, O. Kwan, L. Li, and F.-Y. Wang, "The ChatGPT after: Building knowledge factories for knowledge workers with knowledge automation," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 11, pp. 2041–2044, 2023.
- [2] Y. Tian, X. Li, H. Zhang, C. Zhao, B. Li, X. Wang, X. Wang, and F.-Y. Wang, "VistaGPT: Generative parallel Transformers for vehicles with intelligent systems for transport automation," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 9, pp. 4198–4207, 2023.
- [3] F.-Y. Wang, Q. Miao, X. Li, X. Wang, and Y. Lin, "What does ChatGPT say: The DAO from algorithmic intelligence to linguistic intelligence," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 3, pp. 575–579, 2023.
- [4] Y. Cui, S. Huang, J. Zhong, Z. Liu, Y. Wang, C. Sun, B. Li, X. Wang, and A. Khajepour, "Drivellm: Charting the path toward full autonomous driving with large language models," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 1450–1464, 2024.
- [5] D. Jin and L. Zhang, "Embodied intelligence weaves a better future," *Nature Machine Intelligence*, vol. 2, no. 11, pp. 663–664, 2020.
- [6] A. Turing and J. Haugeland, "Computing machinery and intelligence (p 29-56)," 1950.
- [7] A. Brohan, Y. Chebotar, C. Finn, K. Hausman, A. Herzog, D. Ho, J. Ibarz, A. Irpan, E. Jang, R. Julian *et al.*, "Do as I can, not as I say: Grounding language in robotic affordances," in *Conference on Robot Learning*. PMLR, 2023, pp. 287–318.
- [8] D. Shah, B. Osifski, S. Levine *et al.*, "LM-Nav: Robotic navigation with large pre-trained models of language, vision, and action," in *Conference on Robot Learning*. PMLR, 2023, pp. 492–504.
- [9] X. Li, P. Ye, J. Li, Z. Liu, L. Cao, and F.-Y. Wang, "From features engineering to scenarios engineering for trustworthy AI: I&I, C&C, and V&V," *IEEE Intelligent Systems*, vol. 37, no. 4, pp. 18–26, 2022.
- [10] A. Gupta, S. Savarese, S. Ganguli, and F.-F. Li, "Embodied intelligence via learning and evolution," *Nature Communications*, vol. 12, no. 1, p. 5721, 2021.
- [11] X. Li, Y. Tian, P. Ye, H. Duan, and F.-Y. Wang, "A novel scenarios engineering methodology for foundation models in metaverse," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 4, pp. 2148–2159, 2022.
- [12] Y. Tian, X. Zhang, X. Wang, J. Xu, J. Wang, R. Ai, W. Gu, and W. Ding, "ACF-Net: Asymmetric cascade fusion for 3D detection with lidar point clouds and images," *IEEE Transactions on Intelligent Vehicles*, pp. 1–12, 2023, doi:10.1109/TIV.2023.3341223.
- [13] F.-Y. Wang, J. Yang, X. Wang, J. Li, and Q.-L. Han, "Chat with ChatGPT on Industry 5.0: Learning and decision-making for intelligent industries," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 4, pp. 831–834, 2023.
- [14] F.-Y. Wang, "Parallel system methods for management and control of complex systems," *Decis. Control*, vol. 23, no. 5, pp. 74–76, 2004.
- [15] F.-Y. Wang, "Shadow systems: A new concept for nested and embedded co-simulation for intelligent systems," Tucson, Arizona, USA: University of Arizona, 1994.
- [16] Y. Shen, Y. Liu, Y. Tian, and X. Na, "Parallel sensing in metaverses: Virtual-real interactive smart systems for "6s" sensing," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 12, pp. 2047–2054, 2022.
- [17] Y. Liu, Y. Shen, L. Fan, Y. Tian, Y. Ai, B. Tian, Z. Liu, and F.-Y. Wang, "Parallel radars: from digital twins to digital intelligence for smart radar systems," *Sensors*, vol. 22, no. 24, p. 9930, 2022.
- [18] X. Li, K. Wang, X. Gu, F. Deng, and F.-Y. Wang, "Paralleleye pipeline: An effective method to synthesize images for improving the visual intelligence of intelligent vehicles," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 9, pp. 5545–5556, 2023.
- [19] Q. Wei, H. Li, and F.-Y. Wang, "Parallel control for continuous-time linear systems: A case study," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 4, pp. 919–928, 2020.
- [20] J. Lu, Q. Wei, and F.-Y. Wang, "Parallel control for optimal tracking via adaptive dynamic programming," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 6, pp. 1662–1674, 2020.
- [21] X. Wang, J. Yang, J. Han, W. Wang, and F.-Y. Wang, "Metaverses and demetaverses: From digital twins in CPS to parallel intelligence in CPSS," *IEEE Intelligent Systems*, vol. 37, no. 4, pp. 97–102, 2022.
- [22] Z. Wang, C. Lv, and F.-Y. Wang, "A new era of intelligent vehicles and intelligent transportation systems: Digital twins and parallel intelligence," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2619–2627, 2023.
- [23] Q. Miao, Y. Lv, M. Huang, X. Wang, and F.-Y. Wang, "Parallel learning: Overview and perspective for computational learning across syn2real and sim2real," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 3, pp. 603–631, 2023.
- [24] L. Li, Y. Lin, N. Zheng, and F.-Y. Wang, "Parallel learning: A perspective and a framework," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 3, pp. 389–395, 2017.
- [25] F.-Y. Wang and Y. Wang, "Digital scientists and parallel sciences: The origin and goal of AI for science and science for AI," *Bulletin of Chinese Academy of Sciences*, vol. 39, no. 1, pp. 27–33, 2024.
- [26] F.-Y. Wang, "The DAO to metacontrol for metasystems in metaverses: The system of parallel control systems for knowledge automation and control intelligence in CPSS," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 11, pp. 1899–1908, 2022.
- [27] F.-Y. Wang, "The metaverse of mind: Perspectives on DeSci for DeEco and DeSoc," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 12, pp. 2043–2046, 2022.
- [28] J. Yang, X. Wang, and Y. Zhao, "Parallel manufacturing for industrial metaverses: A new paradigm in smart manufacturing," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 12, pp. 2063–2070, 2022.
- [29] F.-Y. Wang, "New control paradigm for Industry 5.0: From big models to foundation control and management," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 8, pp. 1643–1646, 2023.
- [30] J. Li, X. Liang, R. Qin, and F. Wang, "From DAO to TAO: Finding the essence of decentralization," in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2023)*, 2023, pp. 1–4.
- [31] J. Li and F.-Y. Wang, "The TAO of blockchain intelligence for intelligent web 3.0," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 12, pp. 2183–2186, 2023.

#### ABOUT THE AUTHOR

**Tianyu Shen** Bio of Tianyu Shen can be found at <https://ieeexplore.ieee.org/author/37086603738>.

**Jinlin Sun** Bio of Jinlin Sun can be found at <https://ieeexplore.ieee.org/author/37086371714>.

**Shihan Kong** Bio of Shihan Kong can be found at <https://ieeexplore.ieee.org/author/37086419106>.

**Yutong Wang** (Member, IEEE) Bio of Yutong Wang can be found at <https://ieeexplore.ieee.org/author/37088758926>.

**Juanjuan Li** (Member, IEEE) Bio of Juanjuan Li can be found at <https://ieeexplore.ieee.org/author/38468375000>.

**Xuan Li** Bio of Xuan Li can be found at <https://ieeexplore.ieee.org/author/37085417532>.

**Fei-Yue Wang** (Fellow, IEEE) Bio of Professor Fei-Yue Wang can be found at <https://ieeexplore.ieee.org/author/37277656000>.