AI for Science: Intelligent Scientific Facilities Transforming Basic Research Postprint

Authors: Yang Xiaokang, Xu Yanyan, Chen Lu, Wang Yunbo, Gao Yue, Tian Jidong, Yu Kai, Jin Yaohui, Mei Hong

Date: 2024-03-27T00:00:00+00:00

Abstract

In recent years, artificial intelligence (AI) has achieved numerous remarkable accomplishments in frontier scientific and technological domains, such as AlphaFold2, intelligent control of nuclear fusion, and COVID-19 drug design, demonstrating that AI for Science is emerging as a new research paradigm. To realize fundamental scientific source innovation and its downstream major technological breakthroughs in the intelligent era, two core issues must be addressed: (1) how to leverage the generality and creativity of new-generation AI (particularly generative AI and large models) to drive the formation of new paradigms; and (2) how to utilize AI to empower and transform traditional scientific facilities. This paper proposes a conceptual framework for constructing intelligent scientific facilities that addresses the dual-level requirements of "highly intelligent new scientific facilities"and "AI-enabled existing large-scale scientific facilities," thereby building an AI for Science infrastructure system that creates innovative capabilities including domain-specific large models, generative simulation and inversion, autonomous intelligent unmanned experiments, and large-scale trustworthy scientific collaboration, accelerating major scientific discoveries, transformative material synthesis, and significant engineering technology applications.

Full Text

Abstract

In recent years, artificial intelligence (AI) has achieved remarkable breakthroughs in frontier scientific and technological domains, exemplified by AlphaFold2 for protein structure prediction, intelligent control of nuclear fusion, and AI-accelerated drug design for COVID-19. These advances demonstrate that AI for Science is emerging as a new research paradigm. To realize fundamental scientific innovation and downstream technological breakthroughs in the intelligent era, two core challenges must be addressed: (1) how to leverage the generality and creativity of next-generation AI, particularly generative AI and large models, to foster new paradigms; and (2) how to empower and transform traditional scientific facilities using AI. This paper proposes a conceptual framework for constructing AI-enabled scientific facilities that simultaneously addresses the need for "highly intelligent new scientific facilities" and "AI empowerment of existing large-scale scientific infrastructure." By building a comprehensive AI for Science facility ecosystem, we can enable innovative capabilities including scientific large-scale models, generative simulation and inversion, autonomous intelligent unmanned experiments, and large-scale trustworthy research collaboration, thereby accelerating major scientific discoveries, transformative material synthesis, and critical engineering applications.

Keywords: AI for Science, generative AI, large language models, scientific facilities

1. A New Paradigm for Scientific Research: AI for Science

1.1 Evolution of Scientific Research Paradigms

Artificial intelligence applications in frontier science and technology have already yielded impressive results. In biology, *Science* magazine selected AlphaFold2 as the top breakthrough of 2021 [?]. In materials science, AI has enabled efficient plasma flow control in nuclear fusion tokamak devices [?]. In drug discovery, AI has accelerated COVID-19 therapeutic design [?]. These breakthroughs and broader trends indicate that AI for Science is becoming a new research paradigm.

In 2007, Turing Award laureate Jim Gray described the historical evolution of scientific discovery through "four paradigms" : experimental observation, theoretical derivation, computational simulation, and data-driven discovery (data-intensive science) [Figure 1: see original paper]. Thousands of years ago, humans described natural phenomena through experimental observation and experience, forming the empirical paradigm (first paradigm). Several centuries ago, scientists began using theoretical derivation for research, explaining nature through theories and models, exemplified by Newton' s laws in the 17th century—this became the theoretical paradigm (second paradigm). Since the mid-20th century, scientists have used computer simulations to model complex phenomena and verify theories, establishing the computational paradigm (third paradigm). Over the past decade, the big data era has enabled understanding complex systems through large-scale data analysis, yielding previously unknown scientific theories and forming the data-intensive paradigm (fourth paradigm). However, limited by data acquisition and simulation capabilities, scientific hypotheses in this paradigm remain dominated by expert experience. Moreover, the lack of effective data sharing mechanisms and the localized nature of experiments constrain large-scale, interdisciplinary research activities.

In recent years, deep learning—particularly generative AI—has advanced rapidly, enabling academia to model and mine high-dimensional research data, capture scientific laws underlying multimodal data, and overcome limitations of experimental observations and numerical simulations through data generation. This expands the space of scientific hypotheses. Multimodal large language models offer novel capabilities for literature comprehension, summarization, and experimental protocol generation. Combined with unmanned experimental systems and open scientific data platforms, these technologies can drive scientific research toward a new "platform collaboration" model. Chris Bishop of Microsoft Research Cambridge refers to AI for Science as the fifth paradigm of scientific research—a novel approach using AI and machine conjecture for discovery. Unlike the previous four paradigms, it not only relies on existing data and equations but can also simulate natural phenomena through machine learning, infer unknown laws, enhance research efficiency and accuracy, and explore broader possibility spaces encompassing both problem and solution domains.

1.2 Domestic and International Initiatives for AI for Science

Recognizing AI' s potential to revolutionize science, the UK launched the "AI for Science and Government" program to apply AI to real-world natural and social science problems. The U.S. National Science Foundation (NSF) initiated an AI4Science research program in 2021 to advance scientific discovery in mathematics and physics. In 2022, the French National Center for Scientific Research (CNRS) established the Artificial Intelligence for Science, Science for Artificial Intelligence Center (AISSAI) to promote interdisciplinary exchange and expand AI applications in scientific research. Since 2019, the Chinese Academy of Sciences and its affiliated institutes have conducted AI for Science research across multiple disciplines, including biomedicine, materials science, computational physics, and quantum computing.

In 2017, China' s State Council issued the "New Generation Artificial Intelligence Development Plan," emphasizing the need to "focus on major frontier scientific issues in AI, balance current needs with long-term development, prioritize breakthroughs in fundamental theoretical bottlenecks that may trigger AI paradigm transformations, and promote interdisciplinary integration to provide strong scientific reserves for sustainable AI development and deep application." In March 2023, the Ministry of Science and Technology and the National Natural Science Foundation of China launched a special deployment for "AI-driven scientific research," promoting AI model and algorithm innovation for major scientific problems, developing AI for Science platforms for typical research domains, and gradually building a new model for AI-supported basic and frontier scientific research to accelerate China' s research paradigm transformation and capability enhancement.

2. AI-Enabled Scientific Facilities

While AI has achieved remarkable results in specific scientific domains—such as AlphaFold2 [?], intelligent nuclear fusion control [?], AI-planned automated organic synthesis platforms [?], molecular dynamics simulation [?], COVID-19 drug design [?], and generative materials inverse design [?]—these efforts remain limited to specific research groups addressing specific problems. The scope of disciplines, scale of research scenarios, and reproducibility of research protocols and results are all constrained. A "platform collaboration" model and infrastructure system for AI for Science remain to be established.

To achieve fundamental scientific innovation and major downstream technological breakthroughs in the intelligent era, two core issues must be addressed. First, how can we establish new scientific intelligence infrastructure for the AI for Science paradigm to systematically and holistically unleash the creativity and generality of next-generation AI (especially generative AI and large models) in basic science, enabling functions like spontaneous hypothesis generation, automatic law deduction, autonomous unmanned experiments, and self-driven trustworthy collaboration to facilitate ultra-large-scale, high-iteration scientific exploration? Second, how can we use next-generation AI to empower traditional scientific facilities? Scientific research primarily involves scientists proposing questions and hypotheses, experimental personnel conducting verification, and institutions and publishers disseminating and sharing results and data. Traditional scientific facilities and research paradigms face difficulties in scientific communication, experimental operation, and data sharing [Figure 2: see original paper]. These challenges are particularly acute for sophisticated large-scale scientific facilities with highly complex research environments. Achieving an efficient closed loop among "scientific questions (scientists)—experimental equipment (experimenters)—research data and literature (institutions and intermediaries)" using next-generation AI is essential not only for new facilities but also represents new requirements and opportunities for upgrading existing infrastructure.

In response to these challenges, in April 2023, the AI Institute at Shanghai Jiao Tong University proposed the concept of an "AI-enabled scientific facility" (AISF) at the Pujiang Innovation Forum' s AI for Science special session. The overarching vision targets world-class scientific frontiers and national strategic needs, addressing both "creating highly intelligent new scientific facilities" and "empowering existing large-scale scientific infrastructure" to build a comprehensive AI for Science facility ecosystem that accelerates major scientific discoveries, transformative material synthesis, and critical engineering applications [Figure 2: see original paper].

AI-enabled scientific facilities integrate cutting-edge technologies including generative AI, large language models, big data, and blockchain to create a human-in-the-loop scientific intelligence infrastructure with a three-layer architecture [Figure 3: see original paper]. The foundation layer provides computational support through high-performance computing and computing networks. The scientific model layer constructs cross-disciplinary, cross-modal scientific large-scale models and "AI research assistants." The experimental application layer enables autonomous unmanned experiments and multi-party research collaboration through AI-operated robots and intelligent experimental environments.

Based on this architecture, AI-enabled scientific facilities can deliver four major innovative functionalities beyond traditional paradigms [Figure 4: see original paper]. First, scientific large-scale models enable cross-modal research content generation, literature review synthesis, automatic task decomposition, and experimental protocol generation, constructing an "AI research assistant" system with comprehensive scientific capabilities. Second, generative simulation and inversion provide scientific phenomenon generation for complex fluids, multiphysics fields, and complex material structures, along with AI-accelerated ultra-large-scale simulation capabilities to mitigate the curse of dimensionality and stimulate scientific intuition. Third, high-throughput autonomous unmanned experiments combine automated laboratories with AI models to achieve "wet-dry closed-loop" autonomous experimental verification in synthetic chemistry, synthetic drugs, and materials genomics. Fourth, large-scale trustworthy research collaboration uses blockchain and swarm intelligence technologies to enable on-chain traceability, rights confirmation, sharing, and circulation of scientific models and datasets, accelerating the dissemination of new scientific ideas and methods.

3. Scientific Large-Scale Models

ChatGPT-like conversational large language models represent a disruptive new generation of AI technology. Building upon foundation models and training them with scientific big data can embed scientific knowledge and capabilities. Through "scientist-in-the-loop" reinforcement learning for scientific norm and ethics alignment, we can create research-oriented large models that may establish an interactive, mutually reinforcing, and accelerating innovation mechanism between AI and basic science.

First, AI drives basic science forward. On one hand, AI enhances research speed, accuracy, and knowledge integration capabilities, explores broader scientific hypothesis spaces, and promotes deep interdisciplinary integration and major discoveries. On the other hand, AI serves as a valuable supplement to traditional research paradigms, effectively enhancing experimental observation, theoretical derivation, simulation, and data-driven capabilities.

Second, basic science drives AI development in return. Scientific large-scale models will set technical benchmarks for domain-specific large models. While foundation models primarily target natural human language and common programming languages, scientific AI models must also handle mathematical formulas, physical equations, chemical formulas, material structures, and genetic sequences across disciplines and modalities, serving as a demonstration for other large models. Compared to internet-oriented foundation models, domain-specific scientific large-scale models enable faster training and iteration. Moreover, their user base of researchers and students generates massive scientific questions and answers, accumulating knowledge and guiding new hypothesis generation. Finally, the high energy consumption of large models due to big data and computing demands requires solutions from both supply and demand sides. On the supply side, we must promote energy technology progress to improve energy efficiency in collection, generation, storage, transmission, and usage. On the demand side, we must reduce computing system energy consumption and pursue green, energy-efficient solutions. As energy costs decrease and AI computing architectures improve, both foundation and scientific large-scale models will unleash greater potential.

To construct scientific large-scale models [Figure 5: see original paper], we need to develop four specialized capabilities on top of foundation models and establish corresponding evaluation benchmarks. First, cross-disciplinary and cross-modal unified input capability. While language models have been applied to biomedicine [**?**], materials science, and chemistry for entity recognition, relation extraction, and domain classification, scientific data includes not only text but also formulas, charts, and molecular structures. Unifying these cross-modal inputs for joint modeling of interdisciplinary expertise is a critical challenge. Second, effective invocation of external scientific tools. Despite strong language understanding, generation, and reasoning capabilities, large models can produce plausible but incorrect content. Scientific large-scale models demand higher professionalism and accuracy. A viable solution is using large models as planning and reasoning engines that call different external scientific tools to improve credibility and accuracy. Third, continuous feedback and evolution capability. Just as human scientists improve through experience and feedback, scientific large-scale models can leverage high-quality researcher feedback to enhance domain expertise, scientific knowledge modeling, and inference capabilities, while using experimental feedback to improve hypothesis generation and protocol optimization. Fourth, hallucination elimination. Current models suffer from hallucinations where generated content contradicts real-world facts or user inputs, failing to meet scientific precision requirements. Effectively screening expert knowledge and using high-precision professional knowledge for reinforcement learning is crucial for eliminating scientific hallucinations. Additionally, targeted optimization and ensemble of base models can enhance credibility and accuracy in specific domains. Fifth, evaluation benchmarks for scientific large-scale models. To accurately assess capabilities and promote rapid iteration, we need comprehensive scientific knowledge test benchmarks covering multiple disciplines. These should test cross-modal, cross-disciplinary data understanding and modeling capabilities, as well as the ability to use scientific tools accurately and robustly for complex tasks, and the capacity to refuse generating false and harmful content.

In summary, scientific large-scale models serve as "AI assistants" to support human scientists, requiring interdisciplinary knowledge backgrounds, cross-modal data processing, external tool invocation capabilities, and continuous evolution through feedback and evaluation. It is important to emphasize that scientific hypotheses proposed by these "AI research assistants" are merely suggestions for human scientists that require careful verification before subsequent scientific demonstration or experimental exploration.

4. Generative Simulation and Inversion

Simulation and computer modeling represent important paradigms for "theory-to-phenomenon" deduction. Simulation space bridges hypothesis space and observation space through human sensory intuition [Figure 6: see original paper]. In fields like nuclear physics where laws and theories are well-established, data quality is high, and observation costs are substantial, computer simulation increasingly serves as an effective experimental supplement. However, traditional numerical simulation methods face at least two limitations: slow iterative computation speeds, especially for large-scale problems requiring massive computing power; and incomplete underlying theoretical models for many complex phenomena, necessitating approximations or ignoring complex high-order physical relationships, which may yield results contradicting actual observations.

Generative AI technology promises to overcome these speed and accuracy limitations. By transforming numerical solving problems into data fitting problems using generative neural networks, we can establish efficient mappings from hypothesis space to simulation space, thereby accelerating solutions. Generative rendering technology can generate scientifically plausible phenomena from simulation space to observation space, enabling closed-loop learning across three spaces and driving law inversion [Figure 6: see original paper].

4.1 Mapping Scientific Laws from Hypothesis Space to Simulation Space

For phenomena with relatively complete theoretical models, generative AI can map hypothesis space to simulation space to accelerate equation solving. The key challenge is embedding theoretical models accurately and effectively into neural network training. One approach trains neural networks on simulation data from traditional numerical simulators, indirectly embedding scientific laws [**?**]. A more direct method converts mathematical equations into neural network loss functions, using scientific priors to rapidly converge solutions near theoretical values [**?**]. However, these methods heavily depend on numerical simulator accuracy. For complex phenomena with incomplete theoretical models, training data may deviate from reality due to oversimplified boundary conditions and assumptions, leading to inevitable error accumulation in trained machine learning solvers.

4.2 Closed-Loop Learning Across "Hypothesis-Simulation-Observation" Spaces

For complex phenomena with incomplete theoretical models, generative AI can learn data mappings from simulation space to observation space [**?**], generating signals and information that are statistically consistent with observed distributions and sensorily realistic. This allows scientists to invert simulation space states from actual observation data as "posterior information" for comparison with "prior information" from simulations, enabling correction of existing scientific laws and even discovery of new phenomena by expanding hypothesis space.

Generative AI has been widely validated to dramatically accelerate solving ultra-large-scale scientific computing problems and mitigate the curse of dimensionality. Taking fluid simulation as an example [Figure 6: see original paper], NeuroFluid [?] uses generative AI inversion methods where machine learning solver-based fluid particle simulations drive neural radiance field (NeRF) rendering, mapping physical laws hidden in data from observation space back to simulation space that scientists can understand and control via state parameters, enabling high-precision inference of fluid motion around complex geometries from natural images. However, generative AI' s interpretability and robustness still lack sufficient theoretical guarantees. Future work must focus on embedding complex fluid dynamics, multi-physics fields, and complex material structures into machine learning solvers, and exploring how to further utilize generative AI for reasoning across the three spaces, particularly for problems with incomplete theoretical models, to stimulate scientific intuition and achieve theoretical model improvement or correction.

5. Autonomous Intelligent Unmanned Experiment Systems

Autonomous intelligent unmanned experiment systems aim to integrate AI and robotics with scientific experimentation, standardizing, scaling, and automating experimental workflows to improve efficiency and reproducibility. MIT' s Coley et al. [?] proposed an AI-planned automated chemical synthesis workflow in 2018, and the Jiang Jun team at the University of Science and Technology of China [?] proposed a literature-reading-based robotic automated synthesis system in 2023. Building on these foundations, AI-enabled scientific facilities emphasize creating an "open collaboration, human-in-the-loop" human-machine fusion experimental model [Figure 7: see original paper]. The unmanned experimental platform and intelligent system workflow includes three steps: (1) automatic protocol optimization, where AI autonomously designs models and optimizes protocols based on scientific hypotheses; (2) autonomous task planning, integrating scientific large-scale models to convert experimental procedures into formal robot operation instructions for complete autonomous planning; and (3) unmanned experimental execution, where robotic platforms execute instructions to conduct autonomous experiments while researchers remotely monitor via human-machine interfaces. Based on individual platforms, we can expand to safe, parallel, collaborative large-scale open experimental platforms.

5.1 Unmanned Experimental Operation Platforms

By operation precision, unmanned experiments operate at micro and macro spatial scales. Micro-scale operations target microscopic particles like living cells and proteins, typically using fixed platforms where the core challenge is increasing high-precision operation throughput. Macro-scale operations emphasize workflow completeness, using mobile robots with manipulators to autonomously move between equipment and complete multi-task, full-process automated experiments. Additionally, humanoid robots can mimic high-precision dexterous manipulation skills more directly in human-machine collaborative environments, promising further improvements in unmanned experimental effectiveness.

5.2 Intelligent Systems for Unmanned Experiments

Beyond hardware platforms, constructing intelligent system software is fundamental for autonomous unmanned experimental platforms. Software drives processes including self-state perception, external environment perception, mobile navigation, instrument localization, experimental operation planning, and control execution. Deep reinforcement learning and imitation learning enable autonomous learning from environmental interaction or expert demonstration trajectories, building mappings between observations and optimal actions [**?**]. Recently, language models like ChatGPT offer new approaches for unmanned experiment intelligence systems, enabling end-to-end mapping of language instructions and observations to robot actions, and automatically decomposing complex tasks into manageable subtasks upon receiving human language commands.

5.3 Multi-Robot Collaborative Open Experimental Platforms

Current automation methods mostly use single robotic arms or mobile robots, lacking long-sequence task scheduling capabilities and exhibiting low experimental throughput. Multi-robot platforms adopt parallel workflows and collaborative scheduling with standardized batch operations, data processing, and disaster recovery protocols to improve efficiency, reduce uncertainty, optimize resource allocation, and enhance flexibility. When platforms cannot autonomously complete complex unseen tasks, researchers can manually guide robots to accomplish macro-micro operations through human-robot collaboration, forming an "open collaboration, human-in-the-loop" fusion model. Experimental result evaluation requires designing appropriate metrics for specific domains, including: (1) success rate consistency with manual experiments; (2) whether operations meet required standards (e.g., precision and positioning accuracy); (3) throughput and efficiency of collaborative platforms; and (4) safety assurance. Safety for large-scale open platforms can reference autonomous driving technology, progressing from fully enclosed environments to researcher-present environments where robots move safely, ultimately achieving human-robot collaborative integration with high autonomy.

6. Efficient, Trustworthy, Large-Scale Research Collaboration

Under the new AI for Science paradigm, large-scale cross-domain collaboration has become essential. AI-enabled scientific facilities support data sharing for AI model development and testing, but require measures to protect intellectual property rights and interests. Decentralized Science (DeSci) has gained attention for using Web3 tools including smart contracts and blockchain to address IP issues and promote data sharing. In the AI-enabled scientific facility architecture, blockchain provides the foundation for secure, trustworthy collaboration environments; federated learning addresses data silos in decentralized settings while ensuring security and efficiency; and internet-based collective intelligence integrates different research modules into unified platforms for efficient large-scale collaboration [Figure 8: see original paper].

First, blockchain-based trustworthy computing enables traceability and rights confirmation for contributions in collaborative research, providing an effective method for building trustworthy environments. AI can enhance blockchain predictability and detect vulnerabilities in smart contracts, while blockchain can solve distributed data sharing and training issues for AI models.

Second, federated learning enables decentralized scientific computing through compatible incentive mechanisms for multi-user collaborative model training and inference. By keeping data local and using iterative model aggregation, federated learning satisfies decentralized requirements for scientific data and models in large-scale collaboration.

Third, collective intelligence in cyberspace. Integrating internet of data, group decision-making, and language models, research communities can form large-scale collaborations that break traditional spatiotemporal limitations. Key technologies include: (1) Internet of data based on digital objects, which connects distributed data platforms and supports heterogeneous scientific data interoperability [**?**]; (2) collective intelligent decision-making strategies where human-machine collaboration through "exploration-fusion-feedback" mechanisms [**?**] enhances decision efficiency in open-source communities; and (3) language model-based agents. With knowledge, reasoning, and chain-of-thought capabilities, language models like ChatGPT can serve as agents for autonomous management and scheduling in complex collaborative tasks, enabling machine-machine collaboration for scientific research.

7. Conclusion and Outlook

AI for Science is becoming a new paradigm driving scientific research, attracting significant attention from governments, universities, and research institutions worldwide. This paper proposes constructing AI-enabled scientific facilities, elaborating a three-layer architecture of "computing support—scientific engine—unmanned experiments" and four innovative functionalities: scientific large-scale models, generative simulation and inversion, high-throughput autonomous unmanned experiments, and large-scale trustworthy collaboration. This forms a new generation of highly digitalized and intelligent scientific facilities while empowering existing major scientific infrastructure.

Building AI-enabled scientific facilities will help solve complex scientific problems, promote interdisciplinary innovation, open new scientific frontiers, and drive engineering technology and future industries. In engineering, these facilities can improve simulation and reasoning capabilities for large-scale complex problems, make more accurate predictions for complex scenarios, and enhance reliability and operational efficiency of major engineering equipment. Industrially, they will facilitate transferring basic research outcomes to industry, enable low-cost, high-trust, standardized CRO (Contract Research Organization) collaboration models using AI and blockchain, and establish "risk-sharing, benefit-sharing" incentive mechanisms to improve transformation efficiency and quality of major scientific achievements, thereby supporting future industrial development.

References

[?] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature, 2021, 596: 583-589.

[?] Degrave J, Felici F, Buchli J, et al. Magnetic control of tokamak plasmas through deep reinforcement learning. Nature, 2022, 602: 414-419.

[?] Zhou Y D, Wang F, Tang J, et al. Artificial intelligence in COVID-19 drug repurposing. The Lancet Digital Health, 2020, 2(12): e667-e676.

[?] Coley C W, Thomas III D A, Lummiss J A, et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. Science, 2019, 365: eaax1566.

[?] Zhao H T, Chen W, Huang H, et al. A robotic platform for the synthesis of colloidal nanocrystals. Nature Synthesis, 2023, 2(6): 505-514.

[?] Zhang L, Han J, Wang H, et al. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. Physical Review Letters, 2018, 120(14): 143001.

[?] Yao Z P, Sánchez-Lengeling B, Bobbitt N S, et al. Inverse design of nanoporous crystalline reticular materials with deep generative models. Nature Machine Intelligence, 2021, 3(1): 76-86.

[?] Lee J, Yoon W, Kim S, et al. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 2020, 36(4): 1234-1240.

[?] Li Z Y, Kovachki N, Azizzadenesheli K, et al. Fourier neural operator for parametric partial differential equations//International Conference on Learning Representations. Vienna: International Conference on Learning Representations, 2021.

[?] Raissi M, Yazdani A, Karniadakis G E. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. Science, 2020, 367: 1026-1030.

[?] Cranmer M, Sanchez-Gonzalez A, Battaglia P, et al. Discovering symbolic models from deep learning with inductive biases. Advances in Neural Information Processing Systems, 2020, 33: 17429-17442.

[?] Guan S Y, Deng H Y, Wang Y B, et al. NeuroFluid: Fluid dynamics grounding with particle-driven neural radiance fields//International Conference on Machine Learning. Baltimore: ACM, 2022: 7919-7929.

[?] Ibarz J, Tan J E, Finn C, et al. How to train your robot with deep reinforcement learning: Lessons we have learned. The International Journal of Robotics Research, 2021, 40(4-5): 698-721.

[?] Luo C R, Ma Y, Jing X, et al. Internet of data: A solution for dataspace infrastructure and its technical challenges. Big Data Research, 2023, 9(2): 110-121. (in Chinese)

[?] Zhang W, Mei H. A constructive model for collective intelligence. National Science Review, 2020, 7(8): 1273-1286.

YANG Xiaokang is Professor and Executive Deputy Dean of the Artificial Intelligence Institute at Shanghai Jiao Tong University, Director of the Key Laboratory of Artificial Intelligence of the Ministry of Education at SJTU, and President of the Shanghai Image and Graphics Association. His research focuses on image processing, machine learning, and AI for Science. E-mail: xkyang@sjtu.edu.cn

MEI Hong is Academician of the Chinese Academy of Sciences, Fellow of the World Academy of Sciences for the Advancement of Science in Developing Countries (TWAS), and Foreign Member of Academia Europaea. He is Professor and Director of the Key Laboratory of High Confidence Software Technologies (Peking University) of the Ministry of Education, Director of the Artificial Intelligence Institute at Shanghai Jiao Tong University, President of the Advanced Institute of Big Data in Beijing, and President of the China Computer Federation (CCF). His research covers software engineering and system software. E-mail: meih@pku.edu.cn

*Corresponding author

Note: Figure translations are in progress. See original paper for figures.