Intelligent Scientific Research (AI4R): The Fifth Paradigm of Scientific Research Postprint

Authors: Li Guojie

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

Abstract

The article refers to "intelligent scientific research" (AI4R) as the fifth research paradigm, summarizing its series of characteristics as follows: (1) Artificial intelligence (AI) is fully integrated into scientific, technological, and engineering research, with knowledge automation and intelligence throughout the entire research process; (2) Human-machine intelligence fusion, where emergent machine intelligence becomes a component of scientific research; (3) Effectively addressing combinatorial explosion problems with very high computational complexity; (4) For non-deterministic problems, probabilistic and statistical models play a greater role in scientific research; (5) Interdisciplinary collaboration becomes the mainstream research approach, achieving the integration of the previous four research paradigms; (6) Scientific research increasingly relies on large-scale research platforms characterized by large models, among others. The article points out that the intelligentization of scientific research is a revolution in science and technology, and the opportunities and challenges it brings will profoundly influence the future of China' s scientific and technological development, calling for scientists across all industries to themselves undergo an intelligent transformation.

Full Text

Preamble

Special Topic: Vigorously Promote Scientific Research Paradigm Transformations

Citation Format: Li Guojie. AI4R: The Fifth Scientific Research Paradigm. *Bulletin of Chinese Academy of Sciences*, 2024, 39(1): 1-9, doi: 10.16418/j.issn.1000-3045.20231007002.

Editor' s Note: With the rapid development of big data and artificial intelligence (AI) technologies, humanity is ushering in a new round of scientific and technological revolution and industrial transformation. Recent breakthroughs in deep learning and other technologies have enabled AI to be widely applied in natural sciences and high-tech fields such as mathematics, physics, chemistry, biology, materials science, and pharmaceuticals, yielding remarkable achievements. The rapid advancement of AI provides new opportunities for enhancing the efficiency of scientific research tools and organizational models. Intelligent tools such as AlphaFold2 and ChatGPT have demonstrated capabilities beyond humans in solving complex problems. Trends indicate that AI for Science is becoming a new research paradigm. The intelligent era has arrived, and the transformation of research paradigms and forms is urgent; we must seize the opportunities and respond proactively. To this end, *Bulletin of Chinese Academy of Sciences* has organized a special topic, "Vigorously Promote Scientific Research Paradigm Transformations," guided by Associate Editor-in-Chief, Academician of Chinese Academy of Engineering, and researcher at the Institute of Computing Technology, Chinese Academy of Sciences, Li Guojie.

AI for Research (AI4R): The Fifth Scientific Research Paradigm

Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190

Abstract

This article refers to "AI for Research (AI4R)" as the fifth research paradigm, summarizing its characteristics as follows: (1) Artificial intelligence fully integrates into scientific, technological, and engineering research, enabling knowledge automation and intelligence throughout the entire research process; (2) Human-machine intelligence fusion, where emergent machine intelligence becomes an integral component of research; (3) Effective handling of combinatorial explosion problems with extremely high computational complexity; (4) Addressing non-deterministic problems, where probabilistic and statistical models play a greater role in research; (5) Interdisciplinary collaboration becomes the mainstream research approach, achieving integration of the previous four research paradigms; (6) Research increasingly relies on large-scale platforms characterized by large models. The article argues that the intelligentization of research represents a scientific and technological revolution whose opportunities and challenges will profoundly impact the future of China' s scientific and technological development, calling on scientists across all fields to achieve their own intelligent transformation.

Keywords: AI4R, emergence, combinatorial explosion problems, non-deterministic computing, large scientific models, research platforms

1. AI4R: Conceptual Proposal

Human scientific research activities can be traced back at least to the 6th century BCE in ancient Greece, where thinkers and scientists such as Aristotle and Euclid made important contributions. Modern scientific research began with the scientific revolution of the 16th-17th centuries, with Galileo and Newton as its founding figures. For several centuries before the mid-20th century, there were only two methods of scientific research: experimental research based on observation and induction (the first paradigm) and theoretical research based on scientific hypotheses and logical deduction (the second paradigm). Since the advent of electronic computers, computer simulation of complex phenomena has become the third research method (the third paradigm). Due to data explosion triggered by internet popularization, a data-intensive scientific research approach has emerged in the past two decades (the fourth paradigm).

In January 2007, Turing Award winner Jim Gray, in his final lecture before his death, envisioned the fourth paradigm of scientific research. His report, titled "eScience: A Revolution in Scientific Methods," viewed data-intensive research as a component of eScience, focusing primarily on data management and sharing, but essentially not addressing the role of artificial intelligence technology in research [1]. Since the "big data" boom, data-driven research has gained increasing attention. However, pure data-driven approaches have obvious limitations; model-driven approaches are equally important as data-driven ones, and the two need to be integrated.

The term "scientific paradigm" was first used by Thomas Kuhn [2] in his seminal work *The Structure of Scientific Revolutions*, primarily referring to the views and consensus formed within various disciplines during specific historical periods regarding professional knowledge. This term has now become very popular and its meaning has broadened. The "research paradigm" discussed in this article refers to scientific research approaches from a macro perspective. In recent years, many scholars have begun advocating for a fifth research paradigm. Microsoft Research, which once vigorously promoted the fourth paradigm, has recently been advocating the fifth paradigm and established a new AI4Science research center [^1^]. In November 2019, the author initiated the 667th Xiangshan Science Conference, and subsequently published a review paper titled "Data Science and Computing Intelligence: Concepts, Paradigms, and Opportunities" in the December 2020 issue of *Bulletin of Chinese Academy of Sciences*, which explicitly proposed launching "fifth paradigm" scientific research. The article pointed out that the "fifth paradigm" is not only about traditional scientific discovery but also about exploring and implementing intelligent systems, emphasizing the organic fusion of human brains and computers, and predicting that in 10-20 years, the "fifth paradigm" might gradually become one of the mainstream paradigms of scientific research [3].

It is still difficult to provide a rigorous definition of the fifth research paradigm, but its characteristics have gradually emerged, which can be summarized as follows: (1) Artificial intelligence fully integrates into scientific, technological, and engineering research, enabling knowledge automation and intelligence throughout the entire research process; (2) Human-machine fusion, where emergent machine intelligence becomes an integral component of research, giving rise to tacit knowledge and machine conjecture; (3) Taking complex systems as the main research objects, effectively addressing combinatorial explosion problems with extremely high computational complexity; (4) Addressing non-deterministic problems, where probability and statistical inference play a greater role in research; (5) Interdisciplinary collaboration becomes the mainstream research approach, achieving integration of the previous four research paradigms, particularly the fusion of first-principles-based model-driven and data-driven approaches; (6) Research increasingly relies on large platforms characterized by large models, with scientific research and engineering implementation closely integrated.

Scientists such as Weinan E have translated "AI for Science" as "Scientific Intelligence," a term that has gained popularity and can serve as a reference for naming and translating the fifth research paradigm. However, intelligent research is not limited to basic scientific research; it also includes the intelligentization of technological and engineering research. The "AI for Science" initiative launched by the Ministry of Science and Technology and the National Natural Science Foundation of China is termed "AI-driven scientific research," but when placed alongside names of paradigms such as experimental, theoretical, computer simulation, and data-driven, it appears less concise. Building on the above, this article refers to the fifth research paradigm as "AI for Research" (abbreviated as "AI4R"), which is relatively more concise, broader in scope, and more profound in meaning.

2. AI4R: Successful Cases

Data-driven research approaches are often fast but not precise enough, while theoretical deduction and computation based on first principles are accurate but not fast enough, only capable of handling small-scale scientific problems. In recent years, artificial intelligence technology has been widely applied in scientific research in fields such as biology, materials science, and pharmaceuticals [4]. AI4R can both improve research efficiency and ensure the required precision, becoming a powerful driving force for scientific research [^2^]. There are many successful cases of AI4R. This article introduces three cases related to the Institute of Computing Technology, Chinese Academy of Sciences (hereinafter referred to as "ICT").

(1) Protein 3D Structure Prediction. Using deep learning technology to predict protein 3D structures represents a milestone achievement in AI4R. To date, AlphaFold 2 has predicted 214 million protein structures from over 1 million species, covering almost all known proteins on Earth. AlphaFold 2 is not only a disruptive breakthrough in structural biology; more importantly, it has removed obstacles in scientists' understanding of artificial intelligence, illuminating the path forward for AI4R. In the past, even when computer scientists accurately predicted protein 3D structures, these were only considered "dry experiment" results that had to be validated by biologists through "wet experiments" before being accepted. Now biologists can trust AI predictions, representing an epoch-making advancement in the scientific community. Before AlphaFold 2 was released, ICT had already achieved internationally leading scientific results in protein 3D structure prediction.

(2) Molecular Dynamics Simulation. The Deep Potential team, a China-US collaboration, adopted a novel "deep learning-based molecular dynamics simulation" approach, expanding first-principles-accurate molecular dynamics simulations to 100 million atoms and improving computational efficiency by over 1,000 times. This represents the first international achievement combining intelligent supercomputing with physical models, leading scientific computing from traditional computational models toward intelligent supercomputing. Jia Weile, the first author of this paper, currently works at ICT. In 2022, he increased the computational scale of molecular dynamics to 17 billion atoms, improved simulation speed by 7 times, and enabled simulation of 11.2 nanoseconds of physical processes per day, representing another 1-2 orders of magnitude improvement over the 2020 Gordon Bell Prize-winning achievement.

(3) Fully Automated Chip Design. In May 2022, ICT successfully used AI technology to design the world' s first fully automatically generated 32-bit RISC-V CPU— "Enlightenment 1." The design cycle was shortened to 1/1,000 of traditional methods, generating 4 million logic gates in just 5 hours [^3^]. This innovative achievement represents a major breakthrough for AI in complex engineering design, foreshadowing that "AI for Technology" has as bright a future as "AI for Science." CPU design requires accuracy above 99.99999999999% (13 nines!), which cannot be guaranteed by neural network methods, including the recently popular large language models. The Chen Yunji team at ICT invented a new method using Binary Speculation Diagrams (BSD) to represent circuit logic, reducing the descriptive complexity of general Boolean functions from exponential to polynomial level. An important discovery from "Enlightenment 1" is that not only language models based on neural networks possess emergent capabilities; BSD-like decision trees also exhibit emergent functions. This unexpected discovery has sparked hopes for intelligent technologies beyond neural networks, suggesting that as long as models are sufficiently complex, other AI technologies may also exhibit unexpected emergent functions.

3. AI4R: A New Research Paradigm Emerging as Humanity Advances Toward the Intelligent Era

Research paradigms continuously evolve with the progress of human productivity. The agricultural era had only the first paradigm; the industrial era saw the rise of the second paradigm; the information era introduced the third and fourth paradigms. Humanity now stands in the intelligent stage of the information era, advancing toward the intelligent era, and the intelligent research paradigm emerges accordingly.

Computer science and technology have been studied for over 80 years since Turing proposed his computational model in 1936. It is now widely believed that all computers are implementations of the Turing machine, though the Turing model was primarily used to study computational undecidability. In 1943, Mc-

Culloch and Pitts proposed the neural computing model, which is computationally equivalent to the Turing model in terms of computability but may be more valuable for automata theory. Von Neumann [5] once pointed out: "The Turing machine and the neural network model each represent an important research approach: the combinatorial approach and the holistic approach. McCulloch and Pitts made axiomatic definitions of underlying components to obtain very complex combinatorial structures; Turing defined the functions of automata without involving specific components." These two technical routes have been in constant competition. Despite the neural network model being suppressed, relevant scholars never stopped their research. It was not until 2012, when Hinton and colleagues' deep learning method made a stunning breakthrough in the ImageNet image recognition competition, that neural network models suddenly became popular.

The currently popular neural network models have not substantially changed from those proposed by McCulloch and Pitts. Their breakthrough achievements in image recognition, speech recognition, and natural language understanding are mainly due to data volumes increasing by several orders of magnitude and computer computing power also increasing by several orders of magnitude—quantitative change leading to qualitative change. In his work *Theory of Self-Reproducing Automata* [^4^], Von Neumann pointed out: "The central concept of automata theory lies in complexity. Ultra-complex systems will give rise to new principles," and proposed an important concept—the complexity threshold. Systems below the complexity threshold will ruthlessly decay and dissipate, while systems that break through the complexity threshold will continuously evolve through diffusion and mutation at the data level and can do very difficult things [5]. Current neural network models have hundreds of billions or even trillions of parameters, possibly approaching the complexity threshold for handling difficult problems. Neural networks do not implement Turing computation according to deterministic algorithms; their main function is "guess plus verification." The currently popular convolutional neural networks can be used to guess what the next word is. Guessing and computing are two different concepts; machines based on neural networks are more appropriately called "guessing machines" rather than "computers," and their efficiency in solving complex problems far exceeds the Turing model. The neural network model is just one of many AI models; as long as the complexity threshold is crossed, other AI models may also exhibit extraordinary functions. AI4R aims to let various AI technologies shine in scientific research.

After more than 60 years of accumulation and development, AI technology has become a powerful tool for driving scientific research and production under conditions of abundant data and computing power, unleashing unprecedented energy. Although achieving true general artificial intelligence still requires a long journey, there is no doubt that intelligentization has become the main pursuit of our era. Misjudging the era will result in historical dimensional reduction strikes. Missing the opportunity of the era' s transformation will be a historic setback.

4. The Hallmark of AI4R: Machine Emergence and Human-Machine-Physical Intelligence Fusion

The 标志性 event of the fifth research paradigm is that machine conjecture played a key role in both AlphaFold 2' s protein structure prediction and the impressive capabilities later demonstrated by GPT-4, indicating that large-scale machine learning neural networks have already exhibited some degree of cognitive intelligence [^5^]. Although developers cannot fully explain how machine cognitive intelligence is generated, practice has proven that in many applications, machine guesses are correct. The emergence of cognitive intelligence beyond conventional computation and information processing in human-made silicon-based products represents an epoch-making change [^6^].

"Emergence" refers to the phenomenon where individuals in a system follow simple rules, and through local interactions form a whole, some unexpected properties or patterns suddenly appear at the system level—that is, "quantitative changes in a system can lead to qualitative changes in behavior." The formation of life, collective behaviors of ant and bird flocks, human brain intelligence, and many human social behaviors all originate from "emergence." It is often said that the 21st century is the "century of complexity science," and "emergence" is the most concerned topic in complexity science. The Santa Fe Institute began exploring emergent behavior in science and society in 1984, attempting to create a unified complexity theory to explain "emergence," but revealing the mechanism of "emergence" remains an open scientific question [6].

Machines possess "tacit knowledge" that humans cannot clearly explain, which is a tremendous shock to our once-established epistemology. Some scholars believe that computers can only mechanically execute programs written by humans, and so-called machine intelligence is nonsense. However, artificial neural networks composed of hundreds of billions of automatically generated parameters are already complex systems with "cognitive" capabilities. Their emergent capabilities are not directly input by programmers but are acquired through machine learning. Therefore, we should acknowledge that humans have human intelligence while machines have machine intelligence. Human-machine complementarity is one of the main features of the fifth research paradigm. In the future, we should strive to achieve "each displaying their own intelligence, and intelligence coexisting" [7].

The "cognitive ability of machines" mentioned here is different from human cognitive ability, and "machine understanding" is also different from human understanding. So-called "machine understanding" means that if a machine can form certain rules through learning to achieve mapping from a symbol space to a meaning space, it is said to have a certain understanding ability of the symbol space. For example, machine translation may not understand semantics but can "map" Chinese to other languages, even for rarely contacted minor languages. AI weather forecasting models may not understand meteorological theory but can make more accurate forecasts than numerical weather prediction. This may be a novel form of "understanding," a form that enables prediction. Just as we can say airplanes have flight capabilities different from birds without entangling whether machines "understand" the same way humans do, understanding and consciousness have different levels of connotation; having understanding ability does not necessarily mean having self-consciousness. Separating understanding ability from self-consciousness helps reduce people's inexplicable fear of artificial intelligence.

Different scholars have different judgments on whether large models formed by machine learning will have emergent capabilities similar to the human brain. Scholars like Hinton have always firmly believed that although artificial neural network neurons are simple, complex machine learning networks have some degree of similarity to the human brain. It is precisely because of the conviction of a few forward-looking scientists who have silently worked for decades that today' s major breakthroughs in AI technology have been achieved. I once asked ChatGPT and "Wenxin Yiyan" : "Do machines really have intelligence?" ChatGPT answered: "Machines do have their own intelligence." Wenxin Yiyan answered: "The current mainstream view holds that machines do not yet have real intelligence." The machines' answers are related to the creators' intentions in selecting learning content. Perhaps the different understandings of machine intelligence between Chinese and American scholars are the underlying reasons for our lag in large model research and development.

5. The Main Objective of AI4R: Effectively Addressing Intractable Combinatorial Explosion Problems

Traditional science can not only reveal some mysteries of nature but also solve many difficult engineering problems, such as manufacturing large aircraft. A large aircraft has millions of parts. Because we understand the function of each component and comprehend the aerodynamic principles of the entire system, its complexity is within our grasp. However, for the brain, even if we understand every single neuron, we still cannot explain how consciousness and intelligence emerge, because the functions and properties of complex systems are not linear sums of their components [8]. In many fields such as biology, chemistry, materials, and pharmaceuticals, scientific problems have enormous hypothesis spaces. For example, the number of candidate small-molecule drugs is estimated at 10^{60}, and the total number of potentially stable materials could be 10^{180}—screening them one by one is completely infeasible. This is the "combinatorial explosion" we often talk about, which mathematicians call the "curse of dimensionality." We have the key to open the door of science but lack the strength to push open the heavy door. After more than 300 years of scientific exploration, the low-hanging fruit on the tree of knowledge has almost all been picked; the fruits remaining at the treetops are almost all difficult, complex ones. Combinatorial explosion problems that the previous four research paradigms could not solve are the main arena for the fifth paradigm.

The goal of artificial intelligence is not to blindly simulate basic human skills such as speech, vision, and language, but to enable AI to have the same ability as humans to understand and transform the world. The human brain does not contain deterministic algorithms but uses non-deterministic methods such as abstraction, fuzziness, analogy, and approximation to reduce cognitive complexity. Von Neumann [5] long ago predicted: "Information theory includes two major parts: strict information theory and probabilistic information theory. Information theory based on probability and statistics is probably more important for modern computer design." The tremendous progress in machine learning in recent years has mainly resulted from adopting probabilistic and statistical models to model and analyze problems we do not fully understand. Machine learning provides tools for cross-scale modeling, enabling modeling and computation across all physical scales. Through trial and error and adjustment, it continuously improves obtained results, pursuing statistical significance and final result acceptability. The correctness of statistical significance and the strict correctness of deterministic computational programs represent different approaches to solving complex problems.

We call the fifth scientific paradigm intelligent research because only by breaking through the ideological shackles of reductionism and classical computational paradigms and adopting a new intelligent paradigm can we address the uncertainties in input, output, and solution processes. Problem complexity changes with the computational model. What we call NP-hard problems are NP-hard for the Turing computational model. Natural language understanding, pattern recognition, and other NP-hard problems can be effectively solved on large models, indicating that large language models (LLMs) are far more efficient than the Turing computational model for solving such problems. The success of AI4R is essentially not a miracle of large computing power but a victory of changing the computational model.

When solving problems of low complexity, people pursue "white-box models"and emphasize interpretability. However, for extremely complex problems, obtaining "white-box models"is difficult in the short term. Scientific research can be viewed as a process of transforming "black-box models"into "white-box models"—that is, gradually advancing from not understanding a phenomenon or process to fully comprehending its internal mechanisms and principles. AI4R reminds us that for deep learning and similar "black-box models," we must have a certain degree of tolerance for a period. We should adopt the principle that "practice is the sole criterion for testing truth," acknowledge the reasonableness of "black-box models" to a certain extent, and conduct in-depth research based on them.

6. Important Feature of AI4R: Platform-Based Research

Today' s research still relies on the intelligence and imagination of individual scientists and engineers. Curiosity-driven research remains an important component of scientific inquiry, but research work increasingly depends on three essential elements: high-quality data, advanced algorithmic models, and powerful computing capabilities. In recent years, the scale of these three elements has rapidly expanded, with big data, large models, and massive computing power beginning to constitute indispensable large-scale research platforms. Platform-based research has become an important feature of the fifth scientific paradigm.

The emergence of ChatGPT sparked a wave of building large models, with model parameter scales far exceeding past imagination. Large models have indeed exhibited functions and performance that small models lack, but where the limit lies remains undetermined. Large models inevitably require massive computing power, and the enormous electricity needed to train them has raised concerns and prompted the scientific community to explore transformative devices and computing systems that dramatically reduce energy consumption. Large language models have currently gained favor mainly in the industry. Whether they can serve as universal knowledge bases to provide foundational knowledge and common sense for large scientific models, thereby improving the generalization ability of scientific large models, is a major scientific question requiring exploration. AI represented by large models is still in its early stages; current AI computing is equivalent to the vacuum tube computer era of scientific computing, and there is an urgent need for major inventions like transistors and integrated circuits.

The popular saying "massive computing power creates miracles" emphasizes the role of model and data scale and is correct to a certain extent. However, from a theoretical perspective, linearly expanding computing power does not fundamentally help enlarge the scale of solvable NP-hard problems. Simply increasing computing power is not a panacea. If Go were expanded to a 20$\times 20 board, requiring only one additional line on each side compared to the 19\times$19 board, the computing power needed for brute-force search would increase by 10^{18} times. The proportion of game positions searched during Go model training to all possible game positions is an almost infinitesimally small number (10^−150). The algorithm for ICT' s fully automated CPU design compresses an almost infinite search space to 10^6. These successful cases all demonstrate that the real miracle-creator is search space compression, achieved through intelligent algorithms and model optimization! World-renowned computer scientist Professor Li Ming proved from first principles that "understanding is compression; large language models are essentially compression" [^7^].

Currently, China has launched hundreds of machine learning models of various sizes, but if they merely imitate large models with small ones without making great efforts in algorithm optimization, model fine-tuning and alignment, and data cleaning and organization, they will only waste massive computing power and struggle to narrow the gap with foreign counterparts.

There are two opposing predictions in the scientific community regarding the future of large models. Some scientists, represented by OpenAI, believe that simply expanding model and data scale and increasing computing power will likely cause future large models to exhibit new functions not currently present, showing better generality. More scholars believe that large models will not maintain their development speed of the past two years and, like other technologies, will move from explosive growth to saturation. Because at the current growth rate where computing power for training large models doubles every three months, continuing for 10 years would require a 1 trillion-fold increase in computing power—an impossible occurrence. It is still too early to conclude which prediction is correct. Large language models may not be the best path to general artificial intelligence but merely a stage in AI development. However, they have greater practical value than technologies used in the previous two AI waves. China must quickly narrow the gap with foreign counterparts in large model research and industrialization, forge a development path suitable for national conditions, and simultaneously strive to explore new AI approaches different from large models.

7. Important Implementation Path of AI4R: Interdisciplinary Collaboration and Integration of Multiple Research Paradigms

The integration of computing science with different disciplines is driving a digital revolution in science. It is no longer reasonable to pursue single-discipline development in isolation; interdisciplinary integration is one of the important implementation paths of the fifth research paradigm—AI4R. Over the past century, disciplines have become increasingly subdivided. There were about 500 disciplines in 1900, approximately 5,000 in 2000—a tenfold increase in 100 years. If this trend continues, there may be 50,000 disciplines by 2100. China' s education sector is also establishing more and more disciplines. Does this trend run counter to the development trend of disciplinary integration? How to vigorously reform China' s research and education in the process of promoting AI4R deserves high attention.

The large research platforms required by the fifth research paradigm are actually intelligent research infrastructures covering the three essential elements of research: besides shared large scientific models and tool software, they also include massive scientific data, knowledge bases, and, of course, uniformly scheduled computing power. The new research paradigm based on large platforms will reduce the cost of obtaining data, models, and knowledge, enhance the application capabilities of algorithms and models, and accelerate the iteration of new knowledge. McCarthy and Nilsson provided another interpretation of AI: AI = Automation of Intelligence. The automation of knowledge acquisition, processing, and storage also requires large platforms to achieve. Building nationally advanced scientific research infrastructure requires full demonstration and careful planning. Among these, the collaborative coordination between cross-domain large scientific models and vertical domain professional models is an important issue to consider. The history of AI development has proven that ignoring model generalization and reverting to previous expert systems is a hopeless path. However, generality is also a relative concept; humans themselves do not possess absolute generality. Developing AI need not pursue ideal generality as the only goal; we should emphasize using large models to improve efficiency and reduce costs within an industry or field. Truly general artificial intelligence will take at least 20 more years to achieve; for the next 20 years, we should adopt a technical route that emphasizes both general and specialized approaches. The construction of computing power networks must consider both regional "block" needs and industry-specific "strip" characteristics, with different industries forming efficient professional sub-networks for knowledge and resource sharing.

AI has been widely applied to the previous four research paradigms. Whether in automated experimental equipment, computer-aided theoretical analysis, visualized computer simulation, or intelligent data mining, AI technology has played a key role. The fifth research paradigm does not replace the original four paradigms; it only highlights its power where the first four paradigms are powerless. The fifth research paradigm is also not the end of research paradigm evolution; sixth, seventh paradigms may emerge in the future. In the fifth research paradigm, model-driven and data-driven approaches are deeply integrated; "data" and "principles" can be transformed into each other. Empirical "principles" can be extracted from "data," and high-quality data can be simulated from first principles. Problems that need to be solved in various fields now mostly require human-computer interaction, with humans in the loop; embodied intelligence through human-machine fusion will play an increasingly important role.

Another characteristic of the fifth research paradigm is the integration of research and engineering. Building research platforms, screening high-quality data, and pushing large models to their limits all require high-level engineers. Today, the world leaders in AI are not first-class universities or national laboratories but startups like OpenAI and DeepMind. These research teams not only possess cutting-edge, original basic research capabilities but also conduct large-scale system development and engineering, and have the ability to develop technology platforms, research products, and promote commercialization [9]. For China to enter the international first echelon in AI, it needs to concentrate national strengths to build new research teams integrating industry, academia, research, and engineering development.

The intelligentization of research is a scientific and technological revolution. The opportunities and challenges it brings will determine whether China' s scientific and technological development widens or narrows the gap with international advanced levels in the next 20 years. What determines our future is not entirely being "strangled" technically by others but our own ideological obstacles. Two perceptions affect our decision-making: (1) All software executed by computers consists of algorithms pre-programmed by humans, so-called machine intelligence is nonsense; (2) AI may produce risks beyond human control, and we must ensure its results are completely safe and trustworthy before allowing its promotion and use. The first perception mainly comes from within the computer science community; the second may mainly come from government departments.

In fact, the emergence of cognitive intelligence in computers is an epoch-making breakthrough that we cannot ignore. Machine cognition is based on randomness and probability distributions; astonishingly correct predictions and so-called "hallucinations" are two sides of the same coin, complementing each other. If we forcibly decide that AI models are not allowed to hallucinate, their emergent capabilities will also disappear. We must develop AI technology in an environment that coexists with hallucinations, with development and security driving each other in dual wheels.

"AI for Science" is essentially "AI for Scientists." AI scientists and engineers are not the protagonists of intelligent research; scientists in various fields are the protagonists, because intelligent modeling in each field must be completed mainly by scientists in that field. Scientists in various fields must undertake this responsibility and need to achieve intelligent transformation themselves. If scientists do not understand computers or AI, promoting AI4R becomes very difficult. Currently, the main resistance to promoting AI4R comes from scientists themselves, as many still believe that intelligentization does not belong to the scope of their science and that interdisciplinary integration is not orthodox science. Only with the active participation of a broad range of scientists can intelligent research embark on a healthy and rapid development track.

References

1. Pan J F, Zhang X L. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Beijing: Science Press, 2012. (in Chinese)
2. Kuhn T S. *The Structure of Scientific Revolutions*. Beijing: Peking University Press, 2012. (in Chinese)
3. Cheng X Q, Mei H, Zhao W, et al. Data science and computing intelligence: Concept, paradigm, opportunities. *Bulletin of Chinese Academy of Sciences*, 2020, 35(12): 1470-1481. (in Chinese)
4. Wang H C, Fu T F, Du Y Q, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 2023, 620: 47-60.
5. Neumann J. *Theory of Self-Reproducing Automata*. London: University of Illinois Press, 1966.
6. Holland J. *Emergence: From Chaos to Order*. Hangzhou: Zhejiang Education Publishing House, 2022. (in Chinese)
7. Lee E A. *The Coevolution: The Entwined Futures of Humans and Machines*. Beijing: China Citic Press, 2022. (in Chinese)
8. West G. *Scale*. Beijing: China Citic Press, 2018. (in Chinese)
9. Stanley K. *Why Greatness Cannot Be Planned*. Beijing: China Translation Press, 2023. (in Chinese)

LI Guojie
Academician of Chinese Academy of Engineering, Fellow of the World Academy of Sciences for the advancement of science in developing countries (TWAS). Pro-

fessor of the Institute of Computing Technology, Chinese Academy of Sciences (CAS). He serves as Associate Editor-in-Chief of *Bulletin of Chinese Academy of Sciences*. He was the Director of the Institute of Computing Technology, CAS. He mainly engages in research on parallel algorithm, high performance computer, future network, artificial intelligence, big data and technology development strategy.
E-mail: lig@ict.ac.cn

Responsible Editor: Wen Yanjie

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