AI FOR SCIENCE

2025

復旦大學 SAIS

nature
research intelligence

ADVISORY COMMITTEE

Chair

Li Jin	Fudan University

Members (alphabetizing by family name)

Wenbo Bu	Fudan University
Xingao Gong	Fudan University
Ya-Qiu Jin	Fudan University
Huisheng Peng	Fudan University
Yuan Qi	Fudan University, Shanghai Academy of AI for Science
Weixiao Shen	Fudan University
Libo Wu	Fudan University, Shanghai Academy of AI for Science
Renhe Zhang	Fudan University

Research team

Chapter 1

Zenglin Xu	Fudan University, Shanghai Academy of AI for Science
Yuan Cheng	Fudan University, Shanghai Academy of AI for Science
Yanqing Yang	Shanghai Academy of AI for Science
Yan Xu	Shanghai Academy of AI for Science

Chapter 2

Xipeng Qiu	Fudan University
Yanwei Fu	Fudan University
Shouyan Wang	Fudan University
Min Yang	Fudan University
Hong Zou	Fudan University

Chapter 3

Shuai Lu	Fudan University
Lei Shi	Fudan University
Ke Wei	Fudan University
Xuening Zhu	Fudan University
Weiguo Gao	Fudan University
Yingzhou Li	Fudan University
Wei Lin	Fudan University
Wei Yang	Fudan University

Chapter 4

Hongjun Xiang	Fudan University
Minbiao Ji	Fudan University
Zhipan Liu	Fudan University
Fenglei Cao	Shanghai Academy of AI for Science
Yue Gao	Fudan University

Chapter 5

Tianlei Ying	Fudan University
Jintai Yu	Fudan University
Lei Liu	Fudan University
Yuan Cheng	Fudan University, Shanghai Academy of AI for Science
Siyu Zhu	Fudan University, Shanghai Academy of AI for Science
Hanchuan Peng	Fudan University
Shuhua Xu	Fudan University

Chapter 6

Hao Li	Fudan University, Shanghai Academy of AI for Science
Hongliang Zhang	Fudan University
Bin Zhao	Fudan University

Chapter 7

Nan Chi	Fudan University
Feng Xu	Fudan University
Qi Liu	Fudan University
Xuan Zeng	Fudan University
Fan Yang	Fudan University
Yue Gao	Fudan University

Chapter 8

Libo Wu	Fudan University, Shanghai Academy of AI for Science
Shiping Tang	Fudan University
Anning Hu	Fudan University
Baohua Zhou	Fudan University
Xiaole Wu	Fudan University
Xiaoming Fu	Fudan University
Shaoqing Wen	Fudan University
Qingfeng Yang	Fudan University
Weiqi Tang	Fudan University

Chapter 9

Tianlei Ying	Fudan University
Bo Yan	Fudan University

Content support

Jolie Wu	Springer Nature
Jiahui Zhang	Springer Nature
Rebecca Dargie	Springer Nature
John Pickrell	Springer Nature

Data support

Rong Ju	Springer Nature
Jean Huang	Springer Nature
Jiayi Chen	Springer Nature
Vivek Aggarwal	Springer Nature

Project management

Xiaochuang Xu	Fudan University
Yanqing Yang	Shanghai Academy of AI for Science
Sharon Wang	Springer Nature
Scarlett Ding	Springer Nature
Yakira Zhang	Springer Nature

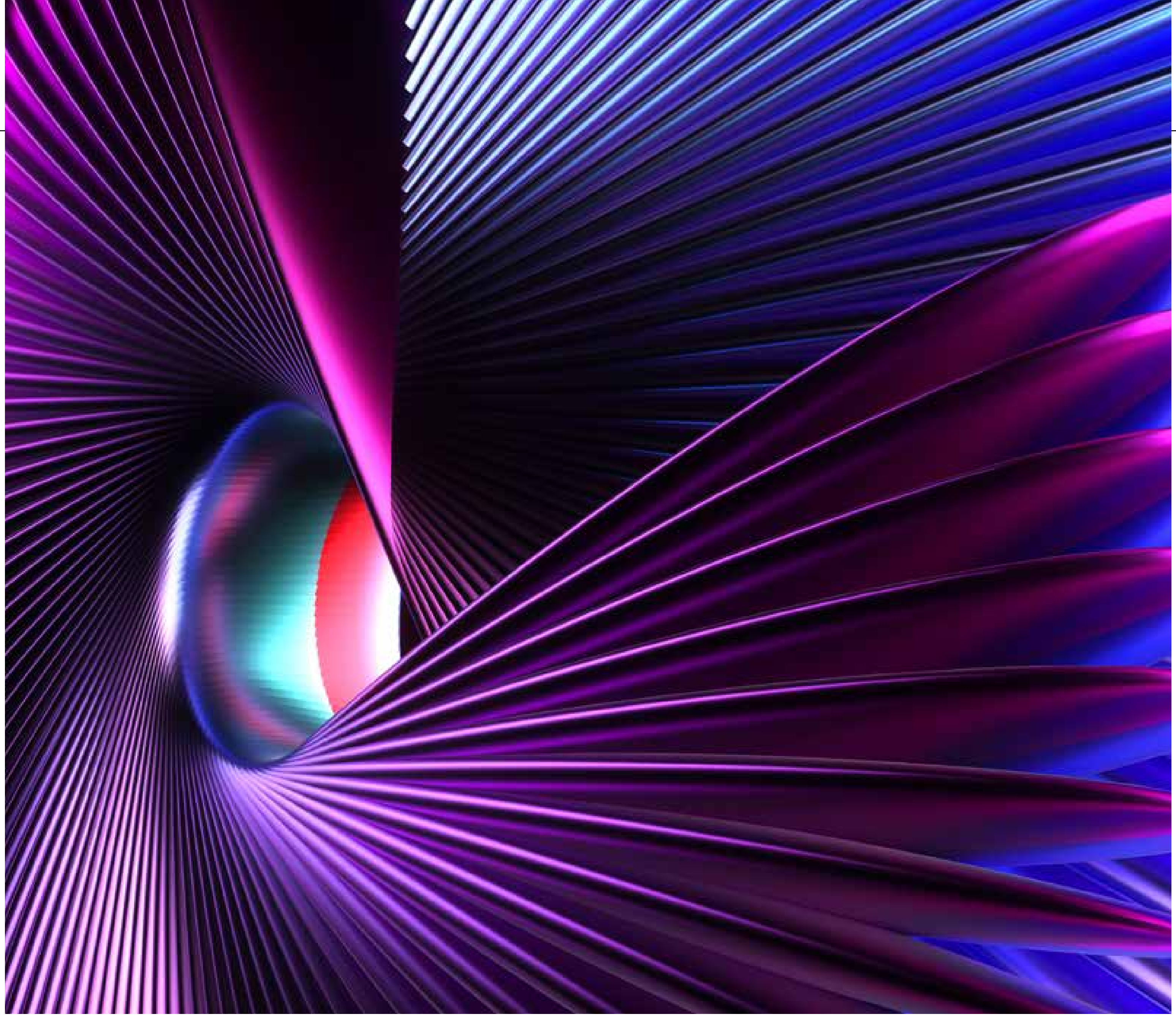
Layout and design

Xinwu Zhao	Springer Nature
Sou Nakamura	Springer Nature

CONTENT

Chapter 1

INTRODUCTION



©diyun Zhu / Moment / Getty

1. The definition and paradigm of AI for Science

1.1 Definition

AI for Science (AI4S) represents the convergence of artificial intelligence (AI) innovation in scientific research and AI-driven scientific discovery, demonstrating their deep integration[1], and the establishment of a transformative research paradigm.

1.2 Paradigm

Scientific research drives advances in AI. Traditional research paradigms can be categorized as empirical induction (experimental science), theoretical modeling (theoretical science), computational simulation (computational science), and data-intensive science[2]. The experimental scientific paradigm generates empirical laws from observations of natural phenomena and reproducible experiments, but does not provide the theoretical foundations that would explain these laws at a fundamental level.

The theoretical paradigm also begins with observations of natural phenomena and reproducible experiments. From these it identifies fundamental scientific problems, and formulates formal hypotheses, and ultimately develops theories through systematic logical reasoning and mathematical analysis. However, verifying these theories within complex systems remains a significant challenge.

Computational science employs numerical methods to simulate complex systems based on scientific models. However, it must simplify these models and requires high-precision computation, inherently limiting fidelity and efficiency.

With technological advances and the exponential growth of data, a new research paradigm of data-intensive science has emerged, using data mining techniques to automatically identify statistical patterns from large-scale datasets, reducing reliance on priori scientific hypotheses. However, it faces limitations in establishing causal relationships, processing noisy or incomplete data, and discovering principles in complex systems.

Modern research confronts complexity challenges, in which interconnected natural, technological, and human systems exhibit multi-scale dynamics across time and space1. Traditional research methods struggle to address these complex challenges effectively, demanding new methods. The need to establish causality has driven the development of innovative inference methodologies capable of handling modern data challenges.

To address the scarcity of high-quality scientific data, such as atmospheric and astronomy data, generative AI technologies such as diffusion models and large language models (LLMs) have been developed. For overcoming limitations in complex system modeling, knowledge-guided deep learning approaches that embed prior knowledge into deep neural networks have been established, significantly enhancing generalization and improving interpretability, such as physics-informed neural networks[3].

AI innovation is reshaping traditional research processes and accelerating discovery. AI integrates data-driven modeling with prior knowledge, which we call model-driven, automating hypothesis generation and validation, enabling autonomous and intelligent experimentation, and promoting cross-disciplinary collaboration. Traditional scientific discovery centres on experimental observations and theoretical modeling, formulates scientific hypotheses and induces general principles, such as physical laws. In contrast, AI employs a model-driven approach to automatically discover hidden patterns from large-scale data, circumventing the need for hypotheses.

Traditional scientific discovery involves generating and validating candidate hypotheses from a large solution space, often characterized by low efficiency and challenges in identifying high-quality solutions[4]. AI harnesses its powerful data processing and analytical capabilities to navigate solution spaces more efficiently, enabling the generation of high-quality candidate hypotheses. For instance, machine learning can assist mathematicians in uncovering new conjectures and theorems[5].

Scientific research depends on the experimental validation of theories. Traditional approaches to experimental design and optimization often rely on manual expertise and iterative trial-and-error processes, which is expensive and inefficient. This is particularly evident in fields such as materials synthesis and fusion experiments.

The integration of AI and robotics can facilitate automated experimental design and execution, leveraging real-time data to refine parameters and optimize both experimental workflows and candidates. AI excels at integrating data and knowledge across fields, breaking down academic barriers and enabling deep interdisciplinary integration to tackle fundamental challenges. This cross-disciplinary collaboration has not only pushed the boundaries of research, but given rise to emerging disciplines, such as computational biology, quantum machine learning, and digital humanities.

References

[1] P. Berens. et al. AI for science: an emerging agenda. *arXiv Preprint*, https://arxiv.org/abs/2303.04217v1 (2023).

[2] T. Hey. et al. The Fourth Paradigm: Data-Intensive Discovery, *Microsoft* (2009).

[3] Raissi, M. et al. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).

[4] Griffin, C. et al. A new golden age of discovery: seizing the AI for science opportunity. *Technical report*, https://storage.googleapis.com/deepmind-media/DeepMind.com/Assets/Docs/a-new-golden-age-of-discovery_nov-2024.pdf (2024).

[5] Davies, A. et al. Advancing mathematics by guiding human intuition with AI. *Nature* **600**, 70-74 (2021).

2. Development and trends

2.1 Recent advances

Breakthroughs in deep learning, generative models and reinforcement learning have enabled AI to identify intricate patterns within vast datasets that are beyond human detection. AI has shown remarkable capacity to autonomously generate scientific hypotheses, designing experimental protocols, and optimizing research pathways.

AlphaFold3[1] has made groundbreaking progress in predicting the structures of almost all types of protein molecules, significantly enhancing the accuracy of protein-ligand interaction predictions, and revolutionizing drug discovery and vaccine design.

Similarly, AI-powered meteorological models such as GraphCast[2], Pangu[3], and Fuxi[4] have dramatically improved global weather forecasting capabilities, enabling longer time-scale and higher-precision predictions.

At the Princeton Plasma Physics Laboratory, reinforcement learning has been applied to optimize plasma control, addressing tearing instability challenges and accelerating the development of nuclear fusion energy[5].

Meanwhile, UC Berkeley and Lawrence Berkeley National Laboratory have established the A-Lab — an autonomous laboratory for the solid-state synthesis of inorganic powders — by combining robotic experimentation with machine learning-driven experimental planning and active learning optimization[6].

2.2 Key challenges and paths

2.2.1 How to build cross-scale AI for science models

Research often requires cross-scale modeling from atomic-level to macroscopic systems. However, current AI models are generally confined to single scales and lack robust mechanisms for effective multi-scale coupling.

The paths to breakthroughs can be explored through the following approaches:

- Embed established physical laws into AI models, enabling cross-scale associations. This approach results in the creation of 'grey-box models' that enhance both the credibility and computational efficiency of models.
- Develop unified neural network architectures that span multiple scales and modalities to enable consistent modeling from micro to macro levels.

2.2.2 How to improve generalization of AI models in scientific research

AI models rely heavily on large-scale training data; however, high-quality data can be scarce. Without sufficient data, models may struggle to learn effective features, limiting their ability to adapt to new fields or tasks, and restricting their application.

Possible paths to breakthroughs:

- Employ generative models to synthesize high-quality scientific data, enhancing sample diversity in data-scarce domains.
- Pre-train cross-domain foundation models and combine them with few-shot learning techniques to enable rapid adaptation to new tasks or disciplinary scenarios.

2.2.3 How to push the boundaries of innovation in AI-assisted scientific discovery

AI is currently constrained to reorganizing and reasoning based on existing knowledge, primarily generating outcomes through pattern recognition and the recombination of existing data. It does not demonstrate genuine creative thinking. Research often necessitates the integration of interdisciplinary knowledge and data, yet AI models face challenges in synthesizing insights from diverse fields. The challenge of enabling AI to actively contribute to the formulation and validation of scientific hypotheses remains an unsolved issue.

Possible paths to breakthroughs:

- Build interdisciplinary knowledge graphs, causal reasoning frameworks, and generative models to integrate multi-domain knowledge bases, enabling AI to extract insights from existing knowledge and propose novel scientific hypotheses.
- Establish a closed-loop system driven by reinforcement learning for AI-assisted experimental design, data analysis, and theoretical modeling, achieving automated scientific discovery.
- Develop visualization tools and interactive interfaces to map AI-generated hypotheses into explainable scientific logical chains, supporting experts in refining and validating theories.

References

1. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
2. Remi Lam et al., Learning skillful medium-range global weather forecasting. *Science* **382**,1416-1421(2023).
3. Bi, K. et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**, 533–538 (2023).
4. Chen, L. et al. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Clim Atmos Sci* **6**, 190 (2023).
5. Seo, J. et al. Avoiding fusion plasma tearing instability with deep reinforcement learning. *Nature* **626**, 746–751 (2024).
6. Nathan, J. S. et al. An autonomous laboratory for the accelerated synthesis of novel materials *Nature* **624**, 86–91 (2024).

3. Data analysis

In this report, AI-related fields are divided into seven broad categories: Core AI (emcompassing algorithms and machine learning), Mathematics, Physical Science, Life Sciences, Earth and Environmental Sciences, Engineering, and Humanities and Social Sciences. The latter six fields are collectively referred to as AI for Science (AI4S). Subsequent sections will follow this classification.

Drawing on data regarding to AI publication volume and citation counts gathered by Nature Research Intelligence, and journals tracked by the Nature Index, we conducted a systematic analysis of global AI-related research from 2015 to 2024. The findings reveal that research in AI and AI4S is undergoing a dual breakthrough in both the scale of publications and a transformative shift in research paradigm.

3.1 A rapid growth in global AI publications, with a surge of AI4S

Between 2015 and 2024, the total volume of academic publications in AI and AI4S expanded quickly. Scientific intelligence emerged as a major driver, accelerating following 2020 and significantly fueling the explosive growth of AI research overall. As illustrated in Figure 1.1, the number of global AI publications nearly tripled over the past decade — soaring from 308,900 to 954,500 — representing an average annual growth rate of 13.7%. The year 2020 marked a pivotal turning point, with the annual growth rate rising from 10.9% before 2020 to 16% thereafter. During this period, the share of publications in Core AI declined from 44.5% to 38.0%, while AI4S gained momentum, increasing its share by 6.4 percentage points — its rapid expansion reflected in the shift from an average annual growth rate of 10.5% before 2020 to 19.3% thereafter. Among the AI4S fields, engineering and life sciences stood out, with their annual growth rates rising from 8.8% and 15.3% (pre-2020) to 16.1% and 28.9% (post-2020), respectively.

FIGURE 1.1 | Total AI Publications Trends and Compositions (2015–2024, in thousands)

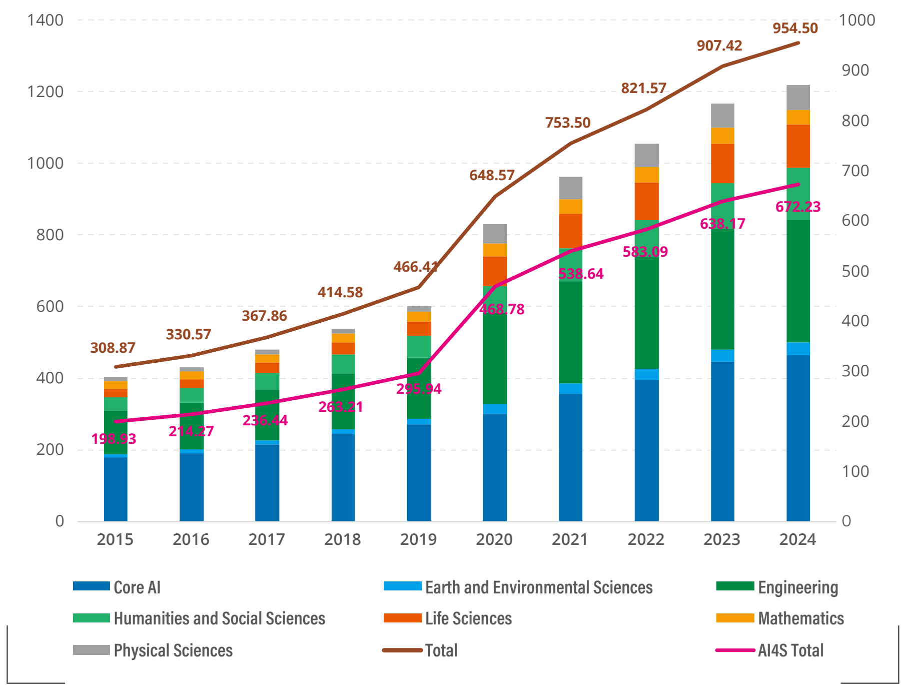


FIGURE 1.2 | Global AI Publications and Top Five Countries/Regions (2015–2024, in thousands)

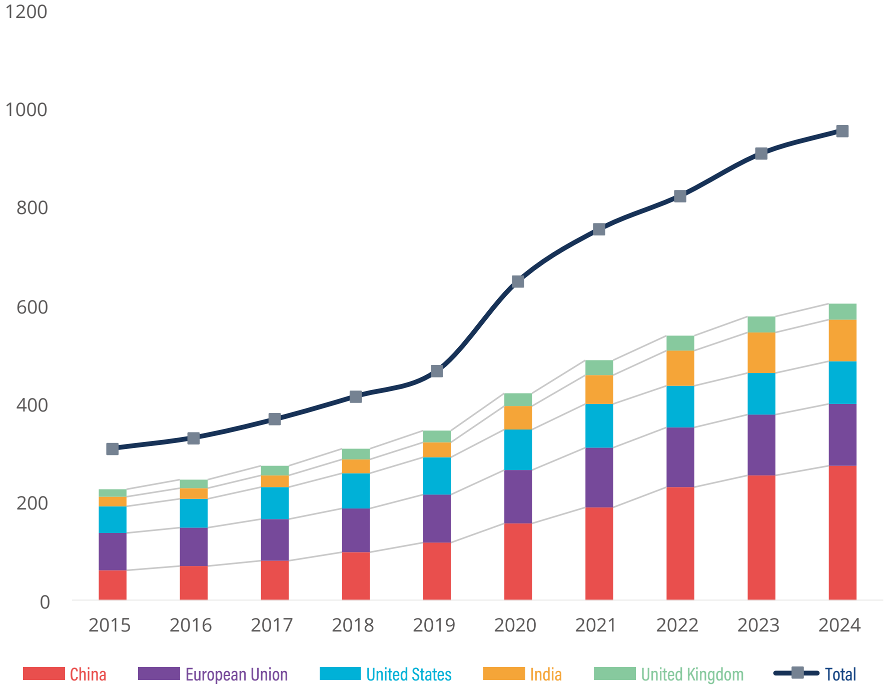


FIGURE 1.3 | Citations of Global AI Publications in Nature Index Journals and Top Five Countries/Regions (2015–2024, in thousands)

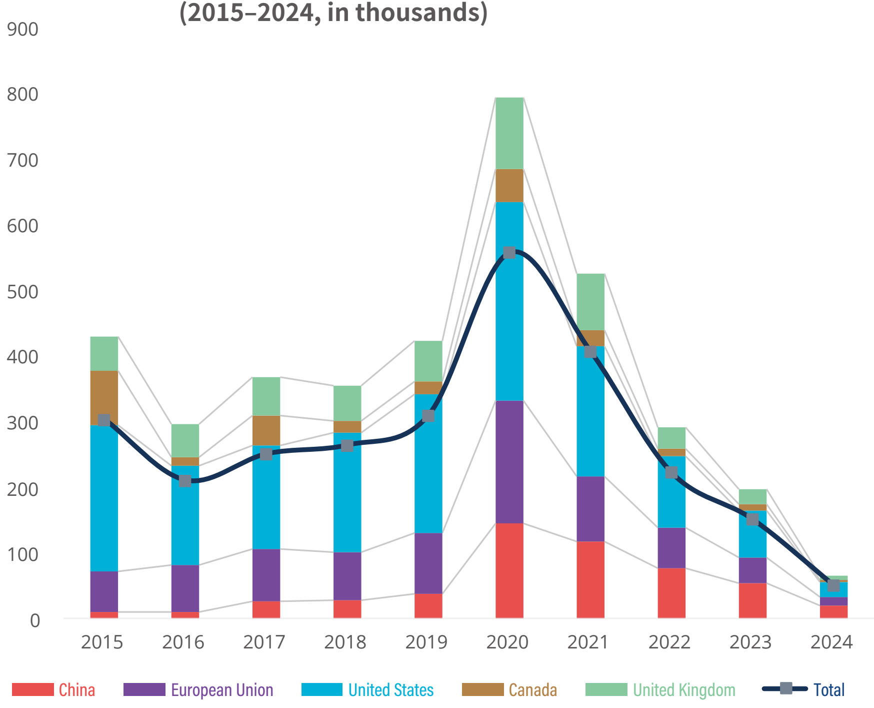
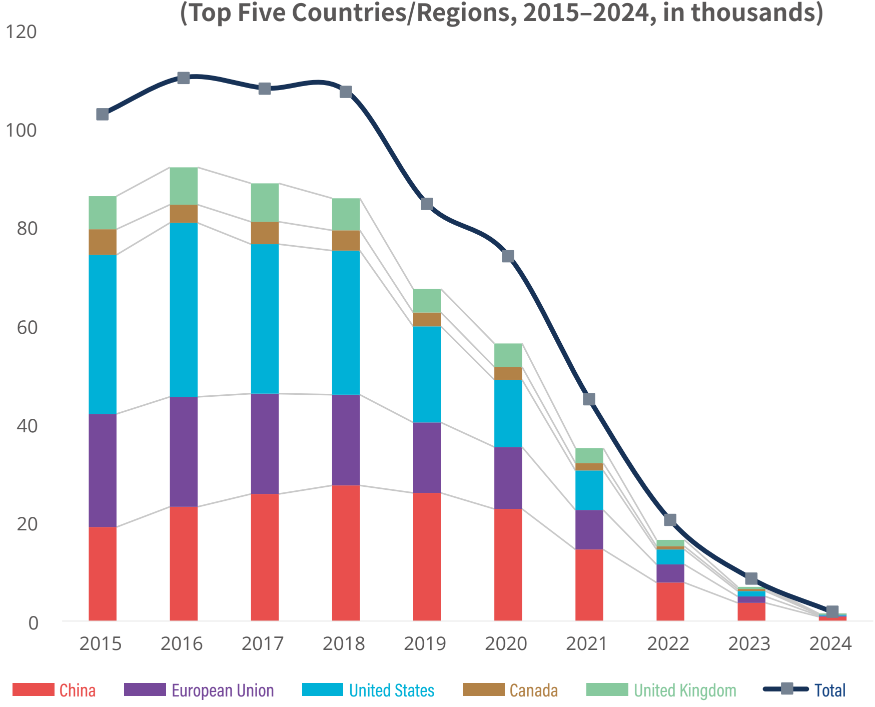


FIGURE 1.4 | Total Citations of AI Publications in Patents, Policy Documents, and Clinical Trials (Top Five Countries/Regions, 2015–2024, in thousands)



Between 2015 and 2024, the global landscape of AI-related publications among the top five countries/regions underwent significant shifts (Figure 1.2). China experienced remarkable growth, with its total number of publications increasing from 60,100 in 2015 to 273,900 in 2024 — accounting for 28.7% of the global total. In 2018, China surpassed the EU in total AI publications to become the world's leading contributor to AI research. By 2022, its output exceeded the combined total of the EU and the US. Meanwhile, India demonstrated a strong upward trajectory, expanding its publications from 18,200 in 2015 — roughly one-third of the US total at the time — to 85,100 in 2024, nearly matching the US output of 85,700 publications.

3.2 The US excels in research quality, while China leads in applied innovation

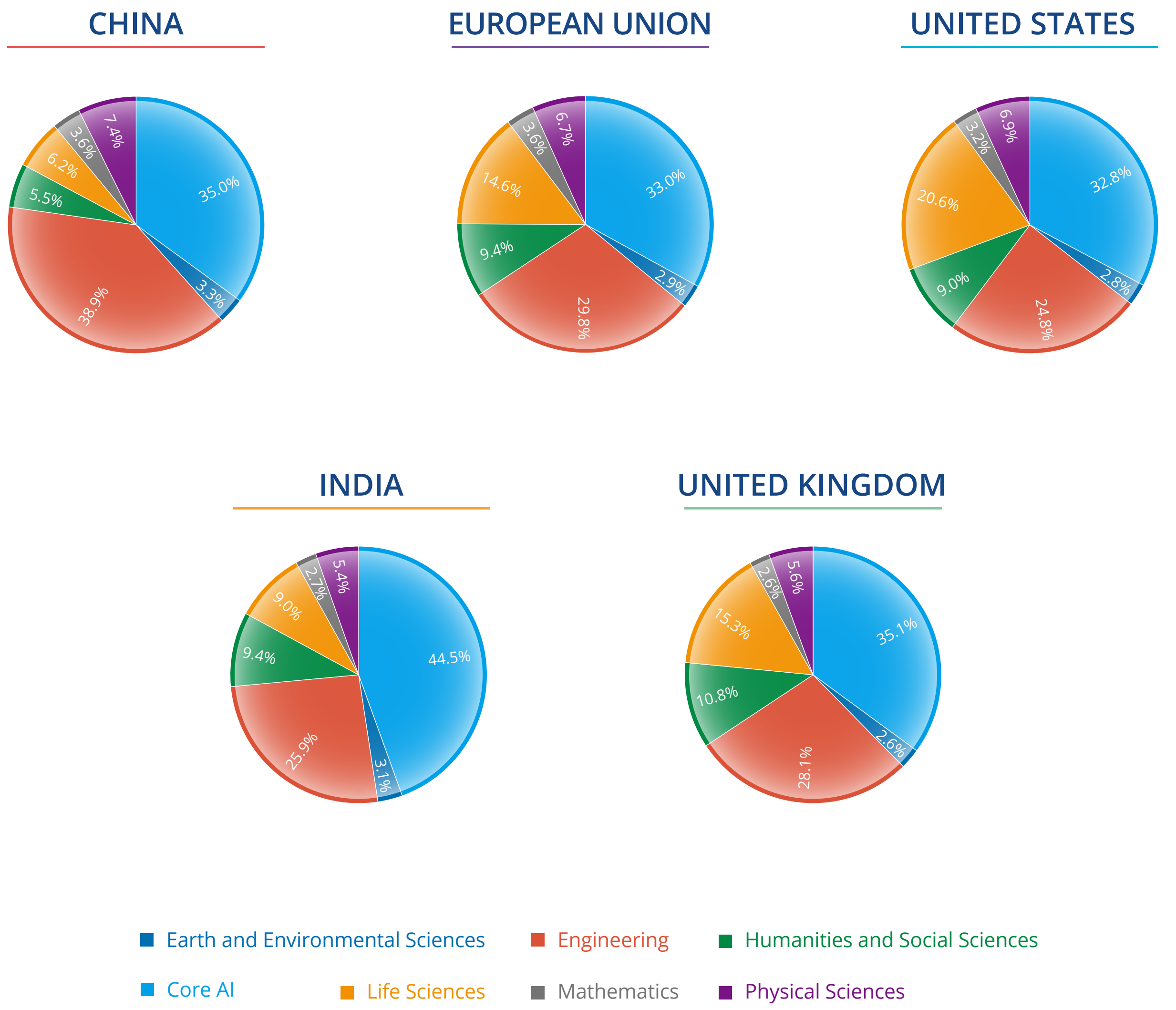
In terms of research quality, the US continues to hold a leading position. As shown in Figure 1.3, citation counts of AI-related papers published in Nature Index journals — a benchmark for high-impact research — place the US consistently at the top, reaching 302,800 citations in 2020. Meanwhile, China's ascent has been more disruptive, with its citations increasing from 10,300 in 2015 to 144,800 in 2020. In 2021, China surpassed the EU for the first time, recording 116,500 citations and securing the second place globally. By 2024, China's share of global AI-related paper citations reached 40.2%, demonstrating rapid progress in closing the gap with the US (which accounted for 42.9% of global citations). It's important to note that since citation data accumulates, the apparent decline in citation trends after 2020 may be distorted; however, this does not significantly affect the overall comparative trajectory among countries.

China has evolved from a follower to a leader in applied AI research. Figure 1.4 highlights citation data from patents, policy documents and clinical trials, illustrating its transformation. With sustained rapid growth, China surpassed the EU in 2016, recording 23,200 citations compared to the EU's 22,300. By 2019, it overtook the US with 26,000 citations, outpacing the US total of 19,600. By 2024, China accounted for 41.6% of all citations of AI publications in patents, policy documents and clinical trials worldwide — cementing its position far ahead of any other country.

3.3 National strengths in AI4S vary, with China and the US remaining dominant research partners

The distribution of AI publications in 2024 highlights distinct national strengths and research priorities within AI4S (Figure 1.5). The US, EU, and the UK focus primarily on engineering (US 24.8%, EU 29.8%, UK 28.1%), life sciences (US 20.6%, EU 14.6%, UK 15.3%), and humanities and social sciences (US 9.0%, EU 9.4%, UK 10.8%). Meanwhile, China's output is predominantly focused on engineering (38.9%), followed by physical sciences (7.4%) and life sciences (6.2%). India's research profile also leans toward engineering (25.9%), followed by humanities and social sciences (9.4%) and life sciences (9.0%).

FIGURE 1.5 | Composition of AI Publications (Top Five Countries/Regions, 2024, in %)



Despite geopolitical competition, global collaboration in AI and AI4S continues to grow steadily. The number of internationally co-authored AI publications surged from 47,200 in 2015 to 133,000 in 2024, while AI4S collaborations increased from 29,900 to 94,800 over the same period — both nearly tripling in volume (Figure 1.6).

China–US collaboration in AI publications peaked in 2020 before experiencing a slight decline, yet it remains the world's largest bilateral scientific partnership. By 2024, China and the US co-authored 12,200 AI publications — nearly double the 6,200 recorded in 2015 (Figure 1.7). Meanwhile, China's scientific collaborations with the EU, the UK, Canada, and Australia strengthened significantly. For instance, China–EU co-authored AI publications surged from 2,200 in 2015 to 9,800 in 2024.

FIGURE 1.6 | Internationally co-authored AI and AI4S Publications (2015–2024)

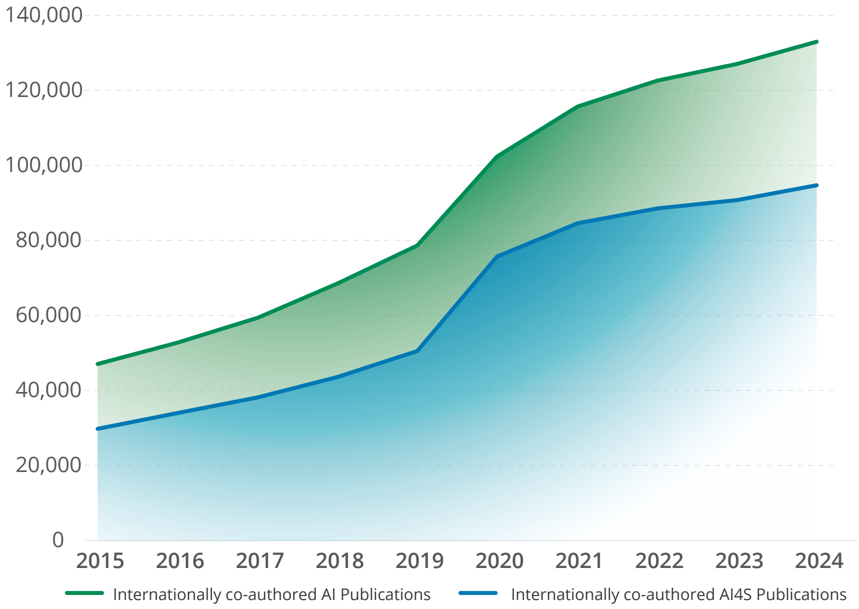
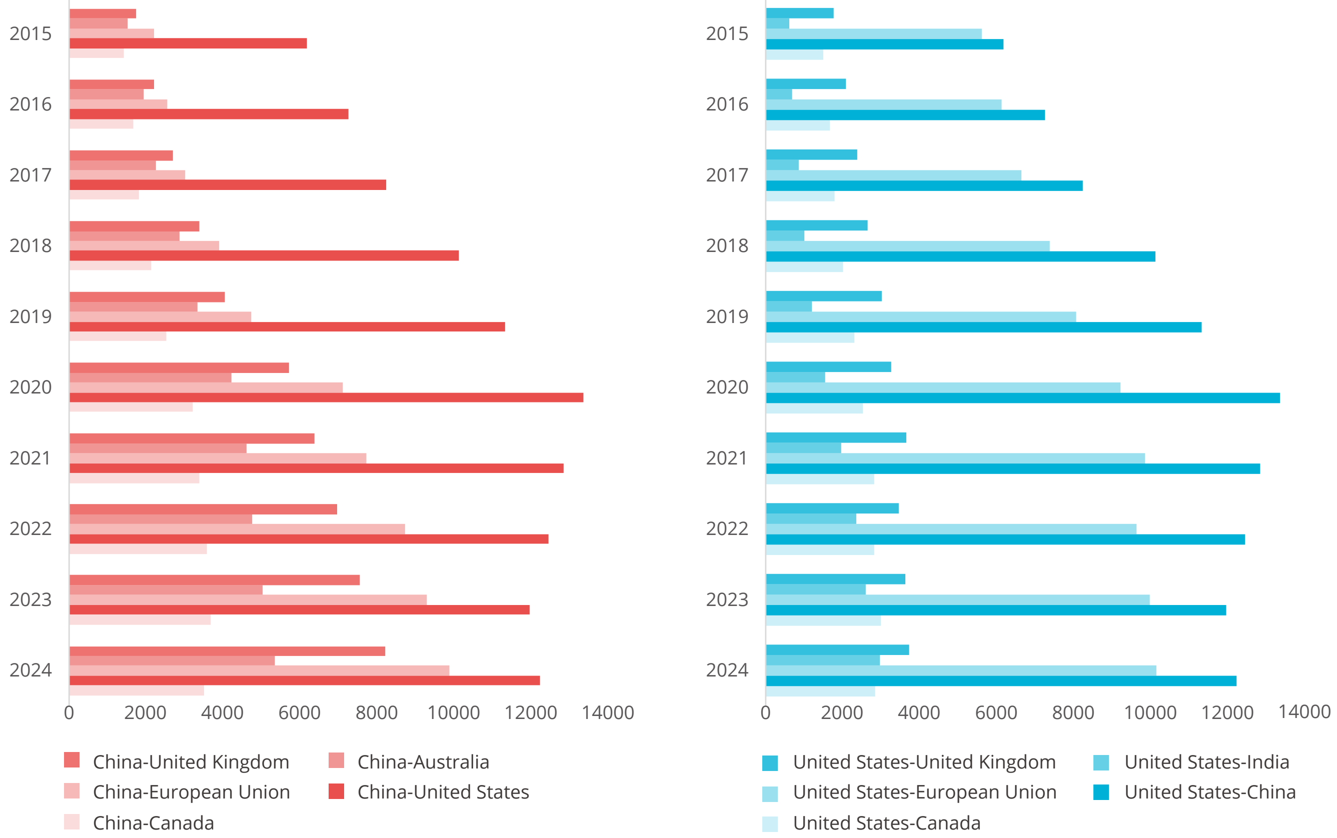


FIGURE 1.7 | International Collaboration in AI Publications: China vs the U.S. (Top Five Collaborations, 2015–2024)



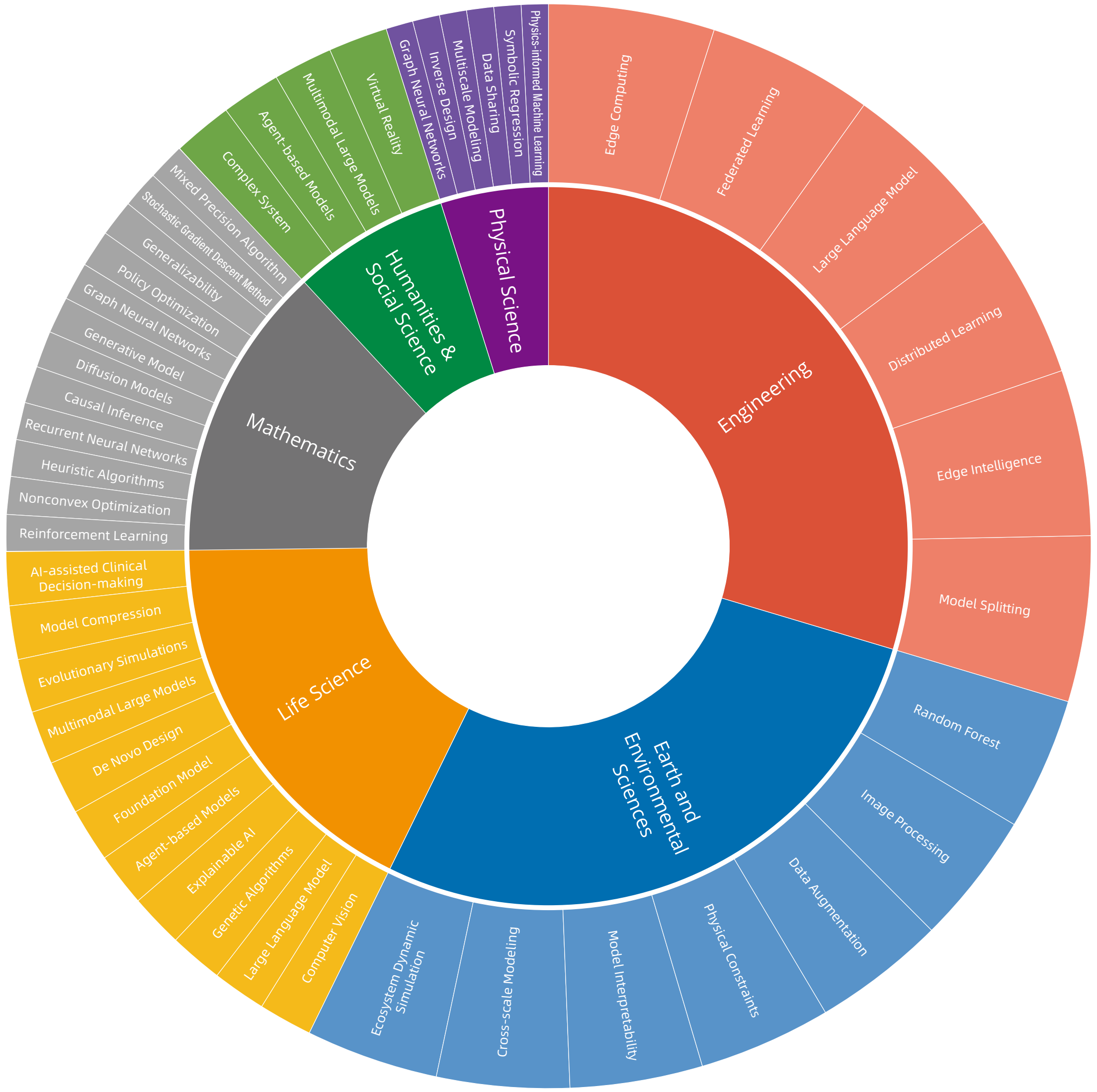
3.4 AI4S is driving a paradigm shift, but what are the most widely adopted AI technologies?

Between 2015 and 2024, AI has sparked an interdisciplinary revolution, marked by the deep integration and adaptation of AI techniques across scientific domains. By aligning scientist-provided keywords with publication databases, we identified and summarized the most widely used AI methods and techniques in AI4S (Figure 1.8).

Today, large language models have emerged as fundamental tools across physical sciences, life sciences, and humanities and social sciences. Reinforcement learning plays a crucial role in complex scenarios, such as engineering system control, mathematical theorem proving, and physical simulations. Meanwhile, computer vision technologies are pervasive in life sciences and earth and environmental sciences. Additionally, distributed learning, graph neural networks, explainable AI and edge intelligence are widely applied across various scientific disciplines.

This evolving AI technology landscape signals a profound transformation: AI is no longer merely an extension or enhancement of scientific tools — it has evolved into a 'meta-technology' that is reshaping the paradigm of scientific discovery. The AI4S-driven revolution is redefining the future of human scientific exploration.

FIGURE 1.8 | AI Technologies Most Widely Adopted in AI4S Research (2015–2024)



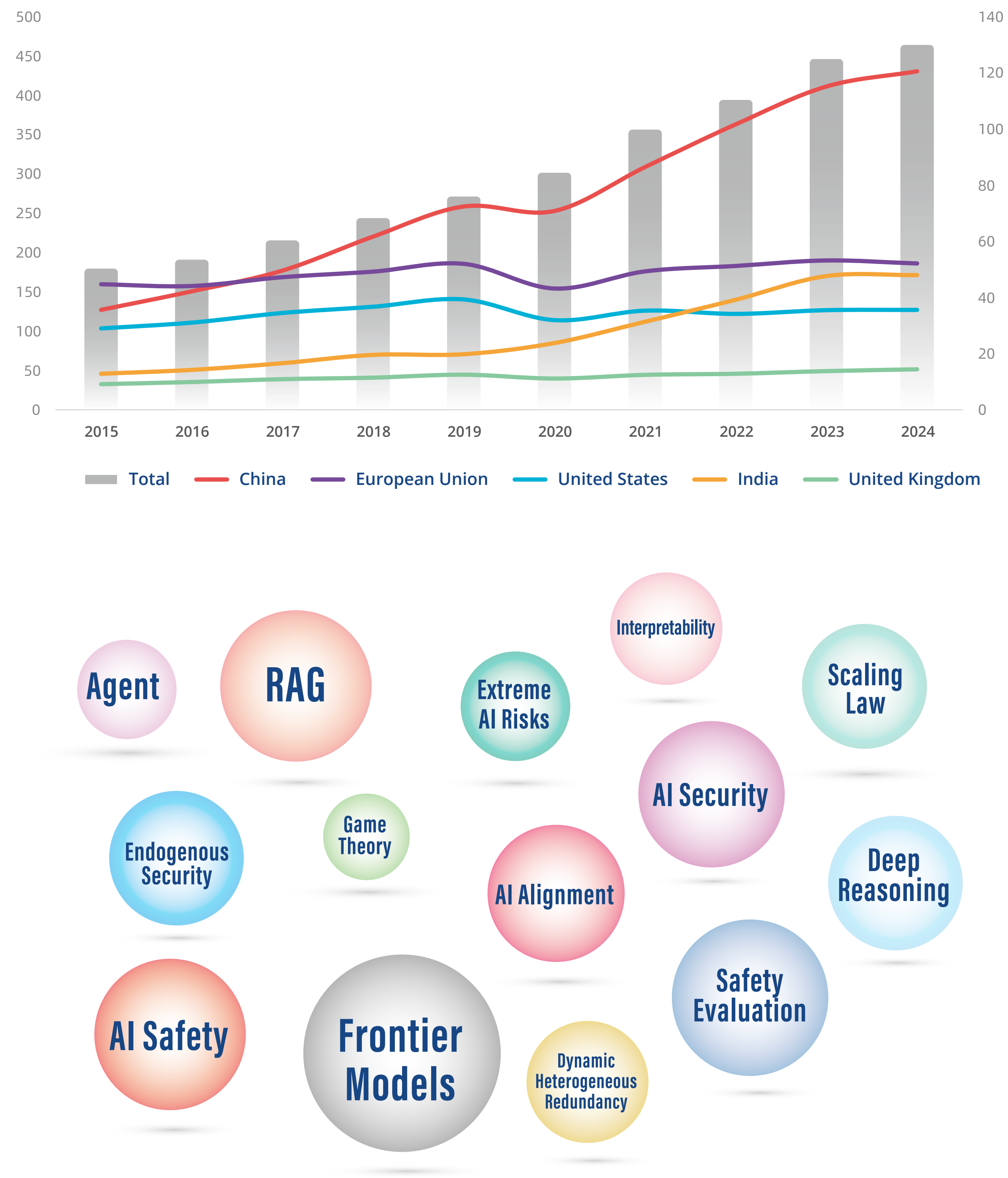
Chapter 2

CORE AI

©Blackjack3D / E+ / Getty

The global volume of Core AI publications has experienced a remarkable surge, rising from 179,200 in 2015 to 463,700 in 2024 (Figure 2). China firmly holds the lead in total publications, while India is rapidly catching up — having surpassed the US and now approaching the EU. The analysis of publication data and the keyword cloud reveals a dual focus in research: sustained innovation in frontier models, foundational algorithms and computing architectures; and, since 2020, a growing focus on AI's endogenous safety, alignment, and extreme risks. This reflects a shifting paradigm that balances breakthroughs in advanced models with the imperative of robust safety and governance.

FIGURE 2 | Core AI – Total Publications, National Trends (in thousands), and Keyword Cloud (2015–2024)



1. From large language models to autonomous agents

1.1 Background

In recent years, large language models (LLMs) [1-3] have driven rapid advances in artificial intelligence (AI), showcasing emergent capabilities that surpass those of earlier technologies. With parameter scales reaching hundreds of billions, these models have achieved remarkable breakthroughs in integrating vast knowledge and enhancing logical reasoning, underscoring their immense potential in the progression toward artificial general intelligence (AGI). Given their immense research and practical value, those working in academia and industry have heightened their focus on LLMs, working to address challenges such as high computational complexity, alignment scalability, and limited interpretability.

At the same time, as training data and computational resources approach their scaling limits, researchers are exploring other laws to further enhance model capabilities. This effort is driving progress in knowledge augmentation[4], multimodal integration[5,6], and deep reasoning[7,8], giving rise to intelligent agent systems capable of autonomous learning and decision-making[9]. This expands the application boundaries of AI, while laying a solid foundation for the pursuit of the true AGI.

1.2 Recent advances

AI technology centered around LLMs is entering a new phase of development. As training data and computational resources approach saturation, researchers have begun exploring the 'second scaling law' — extending beyond the training-phase scaling effects to inference, where innovations in model architecture and hardware-software co-design drive exponential gains in parameter efficiency[10], significantly reducing energy consumption for both training and inference. This strategic shift offers a new pathway for sustained model advances. Key areas of technical progress include:

1.2.1 Knowledge augmentation: To overcome the limitations of LLMs in long-tail and domain-specific knowledge, as well as the challenge of dynamically updating internal knowledge, retrieval-augmented generation (RAG) techniques4 leverage external knowledge bases to enable rapid access to specialized and current information, enhancing the model's reliability and scope of application.

1.2.2 Multimodal integration: Models such as GPT-4o and Gemini[5,6] exemplify multimodal LLMs that integrate visual, speech, and textual information through cross-modal alignment techniques, significantly expanding their practical applications.

1.2.3 Deep reasoning: Models such as OpenAI's o1/o3 and DeepSeek R1[8] have incorporated human-like 'think-reflect' reasoning mechanisms, allowing for extended inference time in exchange for higher-quality responses. These advances have led to remarkable improvements in mathematical problem-solving, scientific analysis, and complex programming tasks.

1.2.4 Autonomous agents: Multi-agent systems[9] harness the cognitive and reasoning capabilities of LLMs, enhancing task execution efficiency and system coordination through autonomous perception, task planning, memory systems, and external tool utilization.

1.2.5 Safety and trustworthiness: Research on the safety and trustworthiness of LLMs is making strides through a three-pronged approach: enhancing interpretability, refining value alignment, and establishing robust evaluation frameworks [11].

These breakthroughs are fueling the surge in AI applications, marking transformative expansion. Industry leaders are integrating LLMs and their derivative technologies into office automation, autonomous driving, intelligent education, and smart healthcare, significantly broadening AI's impact. The ongoing technological revolution and its real-world applications indicate that AGI will continue advancing to higher levels, driving profound transformations across industries and society.

1.3 Key challenges and paths

LLMs must improve in terms of deep reasoning, scaling laws, efficient architectures, full-modality models, emotional cognition, and collective intelligence to overcome critical challenges:

1.3.1 Enhance model reasoning efficiency and generalization: Optimizing reinforcement learning strategies and reward signal design can improve learning and search efficiency. By incorporating human feedback for continuous self-correction, LLMs can overcome complex reasoning and long-sequence generation challenges.

1.3.2 Explore next-generation scaling laws: Beyond scaling laws for pretraining and inference, new expansion principles should be investigated, which involve multi-agent collaboration, real-world physical interactions, and dynamic knowledge updates to sustain model improvements.

1.3.3 Develop high-efficiency model architectures: Integrating distributed training, mixed-precision computing, and specialized hardware acceleration can enhance both training and inference speed, enabling real-time responsiveness and improving overall system efficiency.

1.3.4 Design unified full-modality models: It's essential to drive innovations in architectures that support both understanding and generation across multiple modalities. Addressing challenges such as fine-grained perception deficits, hallucinations, and spatial reasoning limitations will help establish a foundation for modelling.

1.3.5 Advance emotional cognition and adaptive intelligence: Progress in personality simulation and multimodal emotional perception can enhance AI's contextual adaptability, enabling personalized emotion-aware interactions and improving human-computer interaction experiences.

1.3.6 Build collective intelligence through multi-agent collaboration: Research into self-organizing coordination mechanisms for multi-agent integration in complex scenarios will facilitate scalable frameworks for collaborative intelligence, paving the way for human-machine symbiosis.

Accelerating progress toward AGI requires tackling these scientific challenges to drive AI technology forward.

References

1. OpenAI et al. GPT-4 Technical Report. (2023).
2. Touvron, H. et al. LLaMA: Open and efficient foundation language models. *ArXiv preprint* **arXiv:2302.13971** (2023).
3. Sun, T. et al. MOSS: An open conversational large language nodel. *Mach. Intell. Res.* **21,** 888-905 (2024).
4. Lewis, P. et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *ArXiv preprint* **arXiv:2005.11401** (2020).
5. Reid, M. et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *ArXiv preprint* **arXiv:2403.05530** (2024).
6. Zhan, J. et al. AnyGPT: Unified multimodal LLM with discrete sequence modeling. *ACL*, 9637-9662 (2024).
7. Zeng, Z. et al. Scaling of search and learning: A roadmap to reproduce o1 from reinforcement learning perspective. *ArXiv preprint* **arXiv:2412.14135** (2024).
8. DeepSeek-AI et al. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *ArXiv preprint* **arXiv:2501.12948** (2025).
9. Xi, Z. et al. The rise and potential of large language model based agents: A survey. *Sci. China Inf. Sci*. **68,** 121101, (2025).
10. DeepSeek-AI et al. DeepSeek-V3 Technical Report. *ArXiv preprint* **arXiv:2412.19437** (2024).
11. Ma, X. et al. Safety at scale: A comprehensive survey of large model safety. *ArXiv preprint* **arXiv:2502.05206** (2025).

2. Embodied intelligence

2.1 Background

Embodied intelligence refers to intelligent systems based on the perception and action of physical bodies. It interacts with the environment to sense information, plan tasks, make decisions, and generate intelligent behaviour while continuously evolving. Embodied intelligence must be trustworthy — aligning with human values — and be able to self-evolve.

In recent years, significant progress has been made in embodied intelligence research. On a global scale, organizations such as Google DeepMind, OpenAI, Meta, and NVIDIA have achieved numerous breakthroughs in large models, cross-modal perception, and control, such as GPT-4's integration with robotics and RT-X (a high-capacity model) embodied intelligence. Institutions such as MIT, Stanford University, and UC Berkeley are focusing on bionic robots, autonomous operations, and embodied cognition, with research results including Diffusion Policy — a diffusion-based vision-language-action model that optimizes decision-making — and generalizable embodied intelligence frameworks.

In China, institutions and companies such as the Beijing Academy of Artificial Intelligence, Tsinghua University, Peking University, Fudan University, Shanghai Jiao Tong University, and Tencent have made in-depth explorations in multimodal fusion, reinforcement learning, and perception and manipulation skills learning.

2.2 Recent advances

Embodied intelligence technology has advanced remarkably in recent years, driving significant improvements in perception, decision-making, control and commercial applications. High-precision sensors have enhanced the agent's ability to perceive the environment, while the integration of perception and reasoning into dynamic four-dimensional world models has strengthened their understanding and adaptability in complex environments. Additionally, the combination of reinforcement learning and deep learning has significantly enhanced autonomous decision-making and flexible responsiveness.

Recent breakthroughs in cutting-edge research have advanced Visual-Language-Action (VLA) models. Google's RT-2[1] and Physical Intelligence's π_0[2] have demonstrated the capability of generating robot actions directly from visual and language inputs. Helix[3], developed by Figure AI, is the world's first humanoid robot VLA model, featuring an innovative dual-system architecture[4], providing a new paradigm for intelligent robot control. Humanoid robots are evolving rapidly, with Tesla's second-generation Optimus[5] gradually deployed in industrial and service sectors. Multifunctional household robots, such as Stanford University's Mobile ALOHA[6] and NEO Gamma[7] by 1X Technologies, showcase impressive practical capabilities, accelerating the use of robots in everyday life.

Embodied intelligence has also made significant strides in healthcare, neuroscience and virtual reality (VR). For example, the team led by Fumin Jia at Fudan University in Shanghai, China has developed preliminary brain-spine interface technology, partially restoring the ability to walk in several paralysed people. Fudan University's Brain-like Institute has constructed a digital twin platform with billions of neurons[8], driving the development of neuromodulation applications. The integration of VR technology with embodied intelligence enhances immersive experience and drives the future innovation of human-computer interaction. Embodied intelligence is advancing toward industrialization, empowering various sectors such as healthcare, manufacturing, household robotics and VR.

2.3 Key challenges and paths

2.3.1 Basic models

A major scientific hurdle in advancing embodied intelligence is the limited ability of existing foundational models to generalize, which constrains their adaptability across different object categories, scenarios, and tasks. It's critical to enhance their generalization ability across different bodies, scenes, and tasks to achieve universal embodied intelligence.

The development of next-generation foundational models for embodied intelligence is essential, with a focus on exploring Visual-Language Alignment (VLA) and dual-system large models. VLA enhances the robot's ability to understand and execute instructions by aligning visual and language information. The dual-system model consists of a reactive system and a reasoning system, responsible for immediate responses and deep reasoning, respectively, to enhance the robot's decision-making capabilities. These models can be applied in human-robot interaction, and autonomous robot navigation. However, challenges remain in multi-modal learning and system complexity, requiring a balance between efficiency and complexity.

2.3.2 Data engine

The data acquisition and integration for embodied intelligence systems still face challenges such as inconsistent data quality, high data labelling costs, and insufficient cross-modal data synchronization, which limit the generalization and adaptability of the models.

The development of multi-source heterogeneous data acquisition and integration technologies is crucial for building a high-quality data engine. This involves establishing unified data acquisition and multi-modal integration platforms to ensure data standardization and temporal synchronization, while fostering the development of open dataset ecosystems. Multi-modal data fusion technologies — utilizing deep learning and fusion algorithms — integrate information from vision, touch, hearing, and other modalities, enhancing the system's perception and decision-making abilities in complex environments, and improving its adaptability and reliability.

2.3.3 Interaction capabilities

The existing interaction methods for embodied intelligence remain rigid, posing significant challenges in achieving natural and smooth human-machine interactions. The key obstacle lies in enhancing the emotional perception and interaction capabilities of embodied intelligence, enabling it to engage in more natural and human-like exchanges.

Research into affective computing and multi-modal interaction technologies is essential to optimize the interactive experience of embodied agents. By integrating visual, auditory, and tactile information, the immersion and efficiency of interactions can be improved. Given the challenges of real-time responsiveness and stability, the system must be able to respond swiftly while maintaining stable data transmission. When it comes to task planning and execution, it is essential for the agent to have autonomous exploration and decision-making abilities to optimize data acquisition, model generalization, and real-time performance. Balancing personalization and generalization is also essential, as it is important to reduce data dependence and enhance adaptability.

2.3.4 Embodiment development

The existing embodied intelligence hardware still has limitations in flexibility, perception accuracy, and adaptability, hindering its ability to effectively handle complex tasks. A key scientific obstacle lies in how to create a physical carrier that combines agility with adaptability.

It's essential to develop new embodied intelligence hardware that integrates bionic structures, intelligent sensing, and advanced driving technologies. Various sensors will be used to enable environmental perception and self-state monitoring. Visual perception relies on cameras and LiDAR for object recognition and depth perception, while tactile sensors offer force feedback and texture perception. The auditory module processes sound signals through speech recognition. For motion control, path planning, dynamic modeling, and coordinated control methods will be employed to ensure efficient task execution. Flexible electronics and new polymer materials can improve perception accuracy, and enhance data collection capabilities and device performance through integrated design.

2.3.5 Trustworthy mechanisms

The decision transparency and security of embodied intelligence still face significant challenges. Major hurdles revolve around whether intelligent agents act in line with human values, and how to prevent malicious manipulation and data privacy breaches.

A comprehensive trusted evaluation and enhancement system must be established to ensure the reliability of embodied intelligence. It's important to focus research into risk perception and value alignment technologies, ensuring that the behaviour of intelligent agents conforms to ethical standards and societal values. For applications in healthcare, communication, and entertainment, it is crucial to strengthen security mechanisms to prevent risks such as data privacy breaches, malicious manipulation, and unauthorized access. Legal and ethical frameworks must be strengthened to address concerns such as data ownership and accountability, ensuring robust safeguards against ethical dilemmas caused by technological abuse.

2.3.6 Embodied intelligence evaluation

The existing evaluation system for embodied intelligence remains incomplete, without unified benchmarks, making it tough to assess intelligent agents' control, planning ability, and generalization capacity. A scientifically rigorous and structured evaluation system must be developed to advance the field.

To build a systematic embodied intelligence evaluation framework, several critical challenges must be addressed, such as the tendency of existing evaluation methods to focus on a single modality while overlooking the impact of cross-modal integration. The assessment of an agent's generalization ability remains inadequate. The disparity between simulated environments and real-world conditions hampers evaluation accuracy. Synchronization in multi-agent collaborative tasks still requires optimization, and there is an urgent need for a comprehensive evaluation system that rigorously assesses high-level decision-making tasks and low-level execution tasks. Future interdisciplinary collaboration is essential for building a more complete evaluation framework, driving the development and practical application of embodied intelligence technology.

References

1. Brohan et al. RT-2: Vision-language-action models transfer web knowledge to robotic control. *ArXiv preprint* **arXiv:2307.15818** (2023).
2. Physical Intelligence, π_0: A vision-language-action flow model for general robot control. *Technical Report* (2024).
3. Figure AI, Helix: A vision-language-action model for generalist humanoid control (2025).
4. NVIDIA, GR00T N1: An open foundation model for generalist humanoid robots. *Technical Report* (2025).
5. Tesla, Optimus. Available: https://www.tesla. com/AI (2023).
6. Stanford University, Mobile ALOHA: Learning bimanual mobile manipulation with low-cost whole-body teleoperation (2023).
7. 1X Technologies, NEO Gamma. Available: https://www.1x.tech/neo (2025).
8. Lu, W. et al. Imitating and exploring the human brain's resting and task-performing states via brain computing: scaling and architecture. *Nat. Sci. Rev.* **11**, nwae080 (2024).

3. Brain-computer interface

3.1 Background

Brain-computer interface (BCI) technology enables real-time interaction between brain activity and external devices, providing new means to investigate and establish direct causal links between neural signals and brain function or behaviours. BCIs can be used in closed-loop systems that support targeted interventions for brain disorders. Their development also advances AI by improving the decoding of neural signals and inspires new models of brain-like and embodied intelligence. By directly integrating brain signals into intelligent technologies that respond to both the user and their environment, BCIs enable the development of adaptive assistive systems that can flexibly adjust to individual needs and contextual changes — contributing to sustainable, human-centered technologies.

Currently, BCI technology is increasingly integrating neuroscience and AI, evolving from unidirectional neural signal decoding to bidirectional brain-machine interaction and a deeper convergence of biological and artificial cognitive systems.

3.2 Recent advances

Significant progress has been made in neural decoding of motor functions[1]. Recent research explores speech, emotion, and consciousness decoding technologies based on neural activity[2,3]. Biomarkers derived from neural activity have offered critical insights for drug development and precision neuromodulation therapies targeting psychiatric disorders like depression[4]. Neuromodulation techniques employing physical (optical, acoustic, electrical, magnetic) and chemical methods have advanced rapidly[5]. For example, accelerated repetitive transcranial magnetic stimulation (rTMS) brings innovative solutions for depression treatment; focused ultrasound neuromodulation shows promise in Alzheimer's disease treatment; and optogenetics has now entered human clinical research[6].

By integrating neural decoding with neuromodulation or contingent sensory feedback, BCI enables high-precision device control or neural circuit regulation[7]. These advances allow people who are paralysed to restore motor functions[8], and offer new therapeutic strategies for depression[9]. Notably, Medtronic's adaptive closed-loop deep brain stimulation device[10] — now in clinical use in Europe and the U.S. — enhances efficacy for Parkinson's disease treatment through real-time neural circuit monitoring and dynamic modulation, ushering in the era of precision neuromodulation.

Recent AI advances are propelling BCI into cognitive enhancement research, with bidirectional interaction between biological and artificial cognitive systems emerging as a groundbreaking frontier. Scientists have created virtual rodent agents using BCI, accurately predicting neural activity in brains. These virtual agents replicate complex tasks performed by real rodents and even tackle new challenges[11]. This breakthrough not only sets new paradigms for understanding motor control mechanisms, but also signals the emergence of brain-intelligence convergence science, unlocking transformative implications for intelligent robotics, self-learning neuromodulation systems, and brain-inspired AI architectures.

Over 50 years, BCI has evolved from unidirectional signal analysis to bidirectional closed-loop neuromodulation, and is now progressing toward brain-intelligence fusion. The integration of multimodal neuromodulation and decoding algorithms will enable the development of closed-loop perception-decision-modulation systems. The convergence of neural computing and AI drives BCI interaction modalities to expand from neural signal analysis to cognitive interfacing, and integrating physical interfaces with VR and smart environments.

3.3 Key challenges and paths

The human brain is a large-scale complex dynamic system. Its intricate connectivity, time-varying dynamics, and nonlinear interactions pose challenges in decoding and encoding reliability and stability.

How can we achieve specificity regulation of complex neuronal population functions? For excitatory and inhibitory neurons, as well as sensory and motor populations, combining cross-spatiotemporal-scale neural decoding and encoding studies with ultrasound and optical technologies will enable monitoring and modulation systems with millisecond-level temporal precision and micrometre-level spatial resolution. This approach, spanning from single neurons to circuits, offers targeted solutions for neurological and psychiatric disorders.

Precise regulation of neuronal physical-chemical interactions is essential. Brain function is governed not only by electrical activity and circuit dynamics, but also by intricate molecular regulators, such as neurotransmitters, receptor proteins, and ion channels. Integrating molecular networks with neural stimulation technologies can enable multidimensional hierarchical regulation for whole brain function interventions.

To achieve dynamic modulation of brain functions, it's crucial to construct disease-specific biomarker systems with neurophysiological or neurotransmitter data, develop dynamic causal models of neural circuits, and create intelligent BCI chips with integrated sensing, storage, computation, and control functions. This will help to establish closed-loop brain-machine systems for real-time neural monitoring, modeling, and decision-making, enabling the dynamic regulation of nuclei and/or circuits.

It's crucial to integrate biological and AI self-learning strategies. By merging multimodal neural encoding frameworks with embodied AI models through naturalized human-machine interaction paradigms, intelligent agents can enhance their autonomy and environmental adaptability. Dynamic self-learning mechanisms for neural decoding and functional reprogramming will facilitate precise control of robots and devices while establishing virtual agent interaction channels. This will ultimately lead to the creation of immersive brain-machine systems, facilitating real-time interaction between neural and cognitive functions of both biological and artificial brains.

References

1. Hochberg, L. et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* **485**, 372-375 (2012).
2. Chang, EF. Brain-computer interfaces for restoring communication. *N. Engl. J. Med*. **391**, 654-657(2024).
3. Vansteensel, MJ. et al. Fully implanted brain-computer interface in a locked-In patient with ALS. *N. Engl. J. Med*. **375**, 2060-2066 (2016).
4. Wu, W. et al. An electroencephalographic signature predicts antidepressant response in major depression. *Nat. Biotechnol*. **38**, 439-447 (2020).
5. Nasr, K. et al. Breaking the boundaries of interacting with the human brain using adaptive closed-loop stimulation. *Prog. Neurobiol*. **216,** 102311 (2022).
6. Sahel, JA. et al. Partial recovery of visual function in a blind patient after optogenetic therapy. *Nat. Med.* **27**, 1223-1229 (2021).
7. Herron, J. et al. The convergence of neuromodulation and brain–computer interfaces. *Nat. Rev. Bioeng*. **2**, 628-630 (2024).
8. Lorach, H. et al. Walking naturally after spinal cord injury using a brain-spine interface. *Nature* **618**, 126-133(2023).
9. Alagapan, S. et al. Cingulate dynamics track depression recovery with deep brain stimulation. *Nature* **622**, 130-138(2023).
10. Oehrn, CR. et al. Chronic adaptive deep brain stimulation versus conventional stimulation in Parkinson's disease: a blinded randomized feasibility trial. *Nat. Med*. **30**, 3345-3356 (2024).
11. Aldarondo, D. et al. A virtual rodent predicts the structure of neural activity across behaviours. *Nature* **632**, 594-602 (2024).

4. AI system security and safety

4.1 Background

With the emergence of a new wave of generative AI, the challenge of balancing the progress of AI with safety has become a pressing concern[1]. AI systems are vulnerable to security risks across their life cycle, from data collection and inference, to model deployment. Current security frameworks often fall short in addressing these emerging threats. Further, the rapid progress of foundation models empowers frontier AI systems to transform the generated content of foundation models into actions that influence both digital and physical worlds. Without robust safety strategies, the risks could lead to significant consequences. Integrating security considerations into every phase of the development, design, training, and deployment of foundation models and AI systems is essential[2].

4.2 Recent advances

During data collection, adversarial attacks can mislead model training by introducing noise or malicious samples, compromising the integrity of learned patterns. In the training phase, attackers may introduce backdoors which enable covert functionalities. At the inference stage, security threats such as adversarial examples, model hallucinations, and the generation of harmful content pose critical risks. Foundation models, with their vast number of parameters and deep architectures, struggle with interpretability and safety alignment, making them susceptible to threats like jailbreaking and prompt injection[3].

Current approaches centered on 'external hardening' have proven insufficient in addressing the dynamic nature and complexity of AI systems. Furthermore, the field of AI security remains underdeveloped, lacking a fundamental theoretic framework and effective methods for security and safety evaluation. The unpredictable behaviour of foundation models further exacerbates red-line risks associated with frontier AI systems. There is growing recognition that these frontier AI systems have breached critical thresholds, including self-replication4 and scheming[5]. To tackle these challenges, it is essential to develop innovative theories and practices for AI security and safety, prioritizing the integration of internal alignment with external oversight and governance.

4.3 Key challenges and paths

4.3.1 Theory of endogenous AI security

Endogenous AI security requires mechanisms to be integrated into model design, to make safety a fundamental characteristic of AI systems. For example, adaptive mechanisms can dynamically adjust model parameters, and trusted execution environments could safeguard privacy through distributed training. It's crucial to build an endogenous security architecture for AI systems that incorporates trusted computing and a zero-trust framework.

Traditional security paradigms typically rely on passive defense strategies — such as detecting and blocking external threats — which fail to address intrinsic structural flaws that lead to security vulnerabilities. A dynamic heterogeneous redundancy (DHR) architecture shifts the focus from the vulnerabilities in code to the determinism of the system's architecture, allowing for the evolution and dynamic resolution of endogenous security issues at the system level[2]. This complex DHR architecture combines multi-model integration, heterogeneous algorithms, polymorphic execution entities, and strategic scheduling mechanisms, making it challenging for malicious agents to find a stable interface to launch their attacks. By deploying algorithmic diversity, enabling dynamic model migration, and reconstructing feature spaces, the design breaks the static environmental assumptions that attackers rely on, thereby enhancing the stability and security of AI systems in complex environments[7].

4.3.2 Security/safety evaluation and protection of AI Systems

Current evaluation practices primarily rely on static testing and static adversarial attack designs, which makes it challenging to fully evaluate the security and safety of AI systems in open environments. To effectively monitor the security and safety risks associated with foundation models, it is crucial to develop automated, comprehensive, and robust dynamic evaluation technologies.

We need to research automated security and safety evaluation tools which may utilize game theory and machine learning for more adaptive evaluation. Work is also needed on safety and security standards for foundation models and AI systems to reach wide consensus, develop risk-targeted dynamic testing techniques, identify representative and critical risk scenarios, and build a unified AI safety case database. We must investigate risk control mechanisms for the content generated by foundation models, including algorithmic safeguards and content guidelines, to prevent the spread of harmful misinformation.

4.3.3 Proactive discovery and governance on frontier AI system risks

To address the red-line risks associated with frontier AI systems, it is essential to develop proactive risk discovery approaches and model the underlying mechanisms of uncontrolled dangerous capabilities. Establishing a systematic and operable risk evaluation framework is vital for early warning and governance.

Future research should focus on optimizing tool interaction, situational awareness, cognitive reasoning, and memory mechanisms to dynamically model and elicit the potential of foundation models, allowing for the proactive identification of breach points for multiple red-line risks. Additionally, from the perspectives of foundation models, training data, and system software, it is important to develop behaviour control techniques targeting the key dangerous capabilities of these AI systems. It is also important to study effective editing and alignment of model behaviours in relation to identified risks, and design AI system software with innate intelligent risk awareness capabilities.

References

1. Bengio, Y. et al. Managing extreme AI risks amid rapid progress. *Science* **384**, 842-845 (2024).
2. 邬江兴 . 论网络空间内生安全问题及对策 . 中国科学 : *信息科学* **52**, 1929-1937 (2022).
3. Ma, X. et al. Safety at scale: A comprehensive survey of large model safety. *arXiv preprint* **arXiv: 2502.05206** (2025).
4. Pan, X. et al. Frontier AI systems have surpassed the self-replicating red line. *arXiv preprint* **arXiv: 2412.12140** (2024).
5. Meinke, A. et al. Frontier models are capable of in-context scheming. *arXiv preprint* **arXiv: 2412.04984** (2024).
6. 吴铤等 . 基于执行体划分的防御增强型动态异构冗余架构 . *通信学报* **42**, 122-134 (2021).
7. Wei, D. et al. Mimic web application security technology based on dhr architectur. International Conference on Artificial Intelligence and Intelligent Information Processing (AIIIP 2022), *SPIE* **12456**, 118-124 (2022).

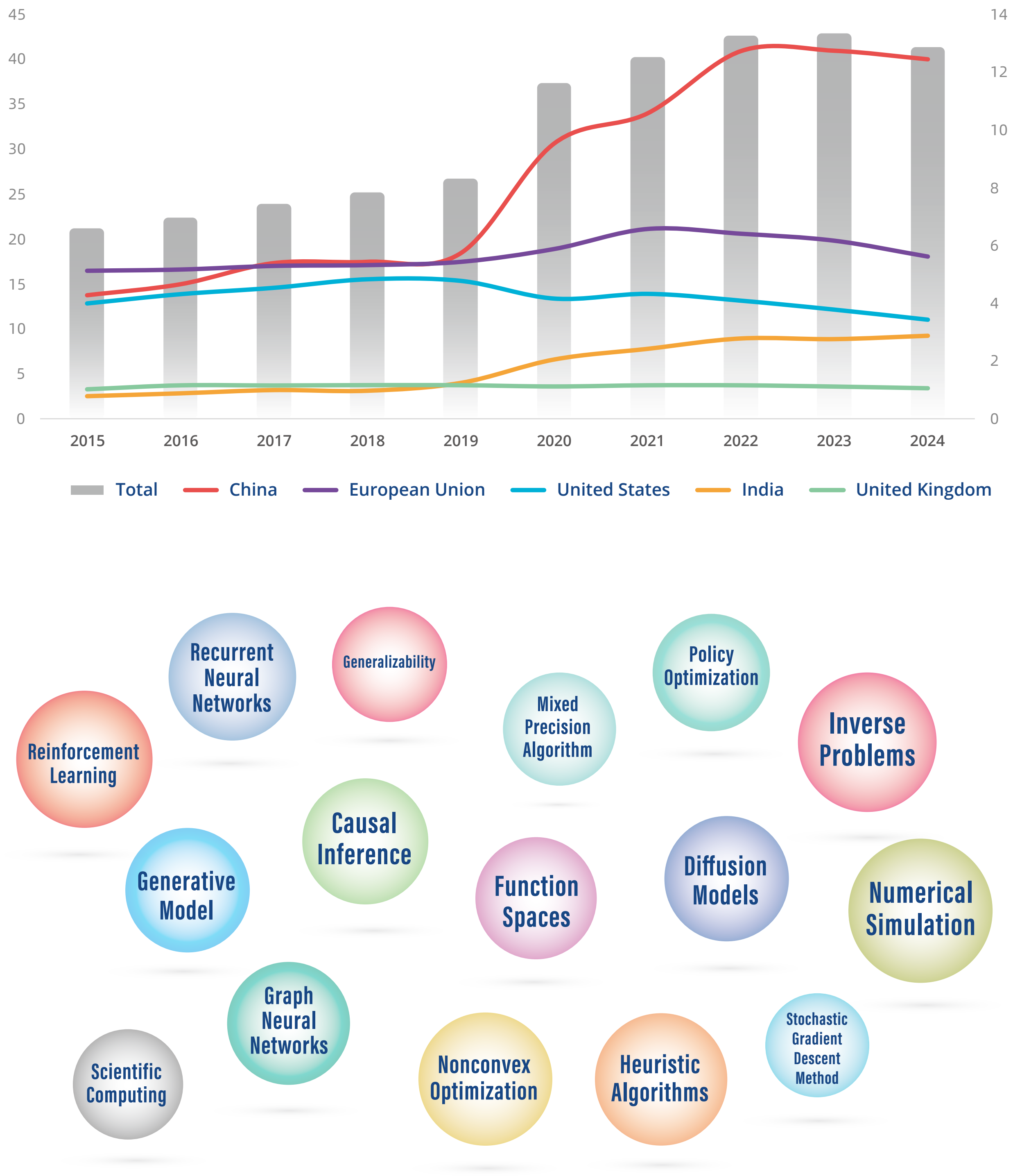
Chapter 3

AI FOR MATHEMATICS

©Constantine johnny / Moment / Getty

Between 2015 and 2024, global AI publications in mathematics grew from 21,200 to 41,200, nearly doubling over the decade (Figure 3). China overtook both the EU and the US in terms of publications produced following 2017, while India experienced accelerated growth after 2020, gradually narrowing the gap with the US. As a cornerstone of AI theory and innovation, mathematics underpins advances in fundamental theory, model design and algorithmic frameworks. In areas such as operations optimization, scientific computing and complex systems, the increasing integration of mathematics and AI is driving collaborative innovation. Data analysis and the keyword cloud highlight the prominence of reinforcement learning, recurrent neural networks, generative models and diffusion models within mathematical sciences, reflecting a paradigm shift toward foundational theory innovation and interdisciplinary integration.

FIGURE 3 | Mathematics – Total AI Publications, National Trends (in thousands), and Keyword Cloud (2015–2024)



1. Mathemathical theory

The development of AI theories can be driven by leveraging mathematical theories and numerical methods. Research in this domain can be divided into three key areas: AI fundamental theory, model design and algorithm implementation.

In AI fundamental theory, one of the key questions is analysing the expressive power of deep learning models. Mathematical tools such as function spaces, approximation theory, and numerical analysis provide methodological support for revealing the intrinsic nonlinear structures within neural networks. These tools lay the theoretical foundation for model stability and generalization capabilities. Researchers use these tools to construct function spaces induced by neural networks, such as Barron spaces and reproducing kernel Hilbert spaces, thereby quantifying network expressiveness. Meanwhile, the universal approximation theorem demonstrates that by widening or deepening feedforward neural networks, they can approximate any continuous function on a compact set to any desired degree of accuracy. In particular, when approximating high-dimensional functions that are essentially low-dimensional, neural networks can effectively avoid the curse of dimensionality, which to some extent explains deep learning's ability to handle complex tasks. Mathematics also plays a crucial role in other network structures. For example, the information propagation and aggregation process in Graph Neural Networks (GNNs) — specialized artificial neural networks that are designed for tasks whose inputs are graphs — can be explained through graph theory and spectral analysis of graph Laplacians. By viewing deep learning models as dynamic systems, the evolution of the hidden states of Recurrent Neural Networks (RNNs) — a class of artificial neural networks designed for processing sequential data — can be analysed using differential equations and stability theory. This not only reveals their long-sequence stability, but also predicts the risk of gradient vanishing that may occur when the activation functions are chosen inappropriately. Equilibrium points, attractors, and bifurcation theory in dynamic systems further provide theoretical support for dynamic behaviour during neural network training, guiding more stable and efficient algorithm design.

In AI model design, mathematical theories can guide network structure design, and the construction of learning paradigms. For example, the architecture of diffusion models is rooted in probability theory and stochastic processes. The forward process uses Markov chains to gradually inject Gaussian noise into data, degrading it, while the reverse process relies on parameterized conditional probabilities to progressively remove noise, achieving reconstruction through a reversible stochastic process. Such design ensures both diverse generated samples and high-quality reconstruction. Additionally, RNNs utilize recursive structures to capture dynamic characteristics of time-series data, with state updates describable by finite difference or differential equations. Residual Networks (ResNets) introduce cross-layer 'shortcuts' to alleviate gradient decay in deep networks, with theoretical analysis depending on linear algebra and differential equations to explain the principle of information identity transmission.

Mathematics is crucial both in data preprocessing and loss function construction and in optimization algorithm design. In data preprocessing, statistical modeling and sample generation techniques can improve data quality. For example, for missing value problems, parameter estimation based on data distribution (such as Gaussian assumptions) and probabilistic interpolation can be used. For class imbalance, oversampling or linear interpolation in feature space can be employed to expand minority class samples, enhancing model robustness. In loss function design, regularization methods — such as L1 regularization promoting sparsity and L2 regularization limiting parameter magnitude — control model complexity and prevent overfitting. Additionally, regularization terms can be designed based on novel function spaces induced by neural networks. Optimization and numerical algorithms rely on mathematical theory, facilitating efficient training processes and supporting model deployment. This includes reducing computational complexity through tensor decomposition or optimizing hardware adaptation via exploring hardware topology.

The key questions in this section can be categorized into two types. The first focuses on revealing the inherent mathematical theories underlying the structure of AI models; the other focuses on the analysis of AI algorithms. The first one involves analysing the expressive power of general simple deep neural network models, followed by analysing the expressive power of complex commonly used models, thereby laying the mathematical foundation for model interpretability and guiding model design based on theoretical analysis. In the second category, research on regularization and implicit regularization theories has laid the foundation for model generalizability, and by further integrating mathematical theories for the target problems into algorithm design, it is expected to obtain trained models with certain generalization capabilities.

Future cutting-edge research should further explore the deep integration of mathematics with AI across various levels, driving the joint progress of theory and practice.

References

1. Barron, A. R. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inform. Theory* **39**, 930-945 (1993).
2. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* **2**, 303-314 (1989).
3. E, W. The dawning of a new era in applied mathematics, *Not. Am. Math. Soc*. **68**, 565-571 (2022).
4. Engl, H. W. et al. Regularization of inverse problems. *Mathematics and its Applications*, **375**, (1996).
5. 5. Pereverzyev, S. V. An introduction to artificial intelligence based on reproducing kernel Hilbert spaces. Springer Nature (2022).
6. 6. Zhou, D. X. Universality of deep convolutional neural networks. *Appl. Comput. Harmon. Anal*. **48**, 787-794 (2020).

2. Optimization

Optimization is one of the key driving forces behind AI and plays a crucial role in every stage of the entire process, including model training, parameter tuning, and performance enhancement. Its techniques are extensively applied across diverse AI scenarios, such as parameter optimization in supervised learning, structure mining in unsupervised learning, maximizing the cumulative reward in reinforcement learning, as well as distributed optimization for large language models.

Most machine learning tasks can be formulated as optimization problems, where algorithms are designed to identify the optimal parameters that are able to achieve good performance. With the rapid development of AI, optimization has been extensively studied. Stochastic gradient descent and its mini-batch version are very fundamental optimization methods, which are suitable for training large-scale datasets. These methods can be further accelerated via the momentum techniques based on historical information, yielding more efficient methods (e.g., AdaGrad and Adam) that may reduce variance and avoid local minima more effectively. More complicated adaptive optimization methods are able to achieve faster and more stable convergence by utilizing second-order information or regularization. These approaches can also incorporate probabilistic statistical methods to mitigate noise, bias, and other uncertainty factors that arise during the training process, enhancing model reliability. Additionally, heuristic algorithms, such as genetic algorithms and particle swarm optimization, have demonstrated strong robustness in solving high-dimensional, non-convex problems. Meanwhile, Bayesian optimization has proven highly effective in various scenarios, such as hyperparameter tuning of deep learning models, automated parameter tuning of machine learning algorithms, and policy optimization in reinforcement learning.

In theories, for convex optimization problems, most gradient-based algorithms are able to converge to the global optimum. As for the more common non-convex problems in machine learning, theoretical guarantees of convergence to stable points or even global minima, can still be achieved under assumptions such as smoothness, weak convexity, or PL conditions. Theories such as optimization dynamics, neural tangent kernels and implicit regularization seek to reveal the underlying principles that enable deep learning to perform effectively in practice and explore the theoretical guarantees for its generalization. When applied to large-scale problems, distributed optimization algorithms must account not only for the computational complexity or iteration complexity, but also for communication complexity.

The development of AI has also significantly advanced the development of optimization, particularly through the powerful generative capabilities of large language models. Traditional optimization modeling relies on domain expertise, but large language models can harness expert knowledge to translate practical problems described by natural languages into structured optimization models and then execute a solver to produce solutions. For example, in delivery route optimization, models can automatically identify key variables, constraints and objective functions, thereby generating mixed-integer programming models, and then by integrating with solvers such as Gurobi and CPLEX, they can generate executable optimization codes to efficiently solve the task. In addition, large-scale solvers often rely on many heuristic rules, but there is no universally acceptable standard for defining them. Traditional approaches typically require manual rule-writing or customizing the search space followed by parameter tuning — both are labour-intensive, and have limited effectiveness. In contrast, large language models can harness existing knowledge to create numerous highly efficient heuristic rules, significantly enhancing algorithmic performance. For example, in the Traveling Salesman Problem and Vehicle Routing Problem, heuristic rules commonly used in genetic algorithms can be written using large language models. Similarly, in satisfiability problems, the heuristic rules in the mainstream CDCL algorithms can also be improved using large language models.

While remarkable progress has been made, the optimization field still faces numerous challenges and opportunities. One of the foremost questions is how to develop efficient algorithms that align with actual computational architecture, an area that remains a constant focus of exploration. In addition to the first-order methods that are mostly used in practice, how to develop efficient second-order algorithms is also an interesting direction to expolore. For certain particular tasks, such as the long sequence, sparse reward policy optimization problem in reinforcement learning, developing effective methods is still challenging. While large language models have begun to show their potential in promoting optimization modeling and algorithm design, further research is required to develop mechanisms that can ensure controllability in their operation. On the theoretical front, studies on the generalization of optimization algorithms in machine learning has made notable progress under ideal conditions, but more significant works are needed before these advances can effectively translate into practical applications.

References

1. Bottou, L. et al. Optimization methods for large-scale machine learning, *SIAM Rev*. **60**, 223-311 (2018).
2. Ahmed, T. et al. Unveiling the potential of Large Language Models in formulating mathematical optimization problems, *INFOR* **62**, 559-572 (2024).
3. Romera-Paredes, B. et al. Mathematical discoveries from program search with large language models, *Nature*, **625**, 468-475 (2024).

3. Statistics

Statistics is the foundation of data science, providing crucial theoretical support and methodological tools to enable AI systems to navigate uncertainty, extract features, and make informed decisions through data analysis, model building and optimization algorithms. The convergence properties and inference in statistics offer a solid theoretical foundation for the interpretability and reliability of AI models. Probability theory and information theory play a decisive role in interpreting model uncertainty and constructing optimization theories, while tools such as linear regression, Generalized Linear Models (GLMs) and high-dimensional data modeling approaches are essential for building models in machine learning and deep learning.

Beyond model building, statistics plays a pivotal role in data preprocessing, feature selection and model evaluation, ensuring the improvement of AI performance. Rapid advances in modern technologies — such as deep learning, generative models and reinforcement learning — has introduced new paradigms and challenges in the integration of statistics and AI.

Deep neural networks, as essential tools in AI, face challenges in statistical convergence analysis. Currently, they are largely empirical, with their internal mechanisms and theoretical proofs often perceived as 'black boxes'. Therefore, a key challenge in this field is using statistical theories to characterize their convergence rates. Nonparametric regression theory can be used to construct convergence rates with least squares or convex losses. At the same time, to ensure robust applicability in real-world scenarios, it is necessary to analyse the convergence performance under more complex conditions, such as non-independent and identically distributed data, time series, and heavy-tailed distributions.

For generative models such as diffusion models and generative adversarial networks (GANs), the theoretical foundation of their statistical properties remains relatively underdeveloped. Two key challenges define research in this area. First, evaluating the effectiveness of generative models in estimating unconditional distributions is crucial. This can be approached by using tools such as the Wasserstein distance and other Hölder-class probability metrics to establish error bounds for distribution estimation, evaluating the effect proof of generator effectiveness.

Although GANs have demonstrated significant advances in learning unconditional distributions for high-dimensional images and natural language analysis, their capabilities in conditional distribution generation remain limited. To tackle this challenge, conditional generators can be explored from two key perspectives. One approach involves conditional interpolation, where researchers can investigate the conditions needed for stable interpolation, ensure the robustness of conditional shift and score function at boundary points, and establish error bounds using the Wasserstein distance and KL divergence. Alternatively, conditional sampling can be employed by appropriately transforming reference distribution samples, aligning joint distributions using KL divergence and performing nonparametric estimation with neural networks.

Reinforcement learning, a pivotal area of AI, also faces substantial challenges in convergence and statistical inference. Currently, it grapples with three critical questions. First, how can statistically effective policy optimization methods be constructed from offline data? A feasible technical approach involves leveraging value enhancement methods to refine performance estimation of reinforcement learning algorithms for a given initial policy. Second, for statistical inference and hypothesis testing in reinforcement learning — such as in business A/B testing scenarios where reinforcement learning is applied for causal inference — it is essential to develop testing methods that accommodate dynamically updated data. Based on Markov Decision Processes (MDPs), which describe the time-varying relationship between treatment and outcome, a sequential testing process is constructed by comparing differences in value functions, thereby analysing the test level and power. Third, how can confidence intervals for policy values be constructed in an infinite-horizon environment? One viable approach is to model the action-value function associated with a policy and leverage its asymptotic normality to construct intervals.

The 'black box' nature of AI models has long been a limiting factor in their interpretability, making statistical causal inference a promising avenue for breakthroughs. Machine learning and deep learning methods often rely on regularization to mitigate variance in high-dimensional estimation, with overfitting sometimes used to counteract the bias introduced by regularization. However, both regularization bias and overfitting can distort estimates, ultimately compromising the accuracy of causal effect inference.

A key frontier in this area is the development of robust debiasing methods for causal inference grounded in statistical theory. By constructing frameworks based on semiparametric theory, and employing Neyman orthogonality and cross-fitting methods for debiased estimation, asymptotic normality can be achieved in causal effect estimates. Moreover, debiased machine learning requires estimating unknown Riesz representations, while the finite sample mean squared error and asymptotic properties of regression estimates remain pressing scientific questions that warrant further exploration.

Statistics plays an fundamental role in AI development, while the rapid advancement of AI technology continues to drive innovation in statistical methods. Complex models, such as deep learning, can automatically capture interaction effects in high-dimensional data through multilayer nonlinear structures. Ensemble learning guarantees robustness in data modeling and prediction. For massive unstructured data, natural language processing techniques can transform texts into semantic embedding vectors, converting them into structured inputs for statistical modelling. Bayesian neural networks introduce probabilistic weights to provide uncertainty quantification for high-dimensional parameter estimation.

As new theories and methods emerge, statistics will continue to efficiently address critical problems in AI, and push AI technology to new heights. A reciprocal relationship — where AI and statistics continually refine and elevate one another — will be instrumental in shaping future intelligent revolutions.

References

1. Huang, J. et al. An error analysis of generative adversarial networks for learning distributions, *J. Mach. Learn. Res.*, **23**, 1-43 (2022).
2. Zhou, X. et al. A deep generative approach to conditional sampling, *J. Am. Stat. Assoc.*, **118**, 1837-1848 (2023).
3. Zhou, Y. et al. Testing for the Markov property in time series via deep conditional generative learning, *J. R. Stat. Soc. B.* **85**, 1204-1222 (2023).
4. Luo, L. et al. Multivariate dynamic mediation analysis under a reinforcement learning framework, *Ann. Stat.* **53**, 400-425 (2025).
5. Shi, C. et al. Statistical inference of the value function for reinforcement learning in infinite-horizon settings, *J. R. Stat. Soc. B.*, **84**, 765-793 (2022).
6. Shi, C. et al. Dynamic causal effects evaluation in a/b testing with a reinforcement learning framework, *J. Am. Stat. Assoc.* **118**, 2059-2071 (2023).

4. Scientific computing

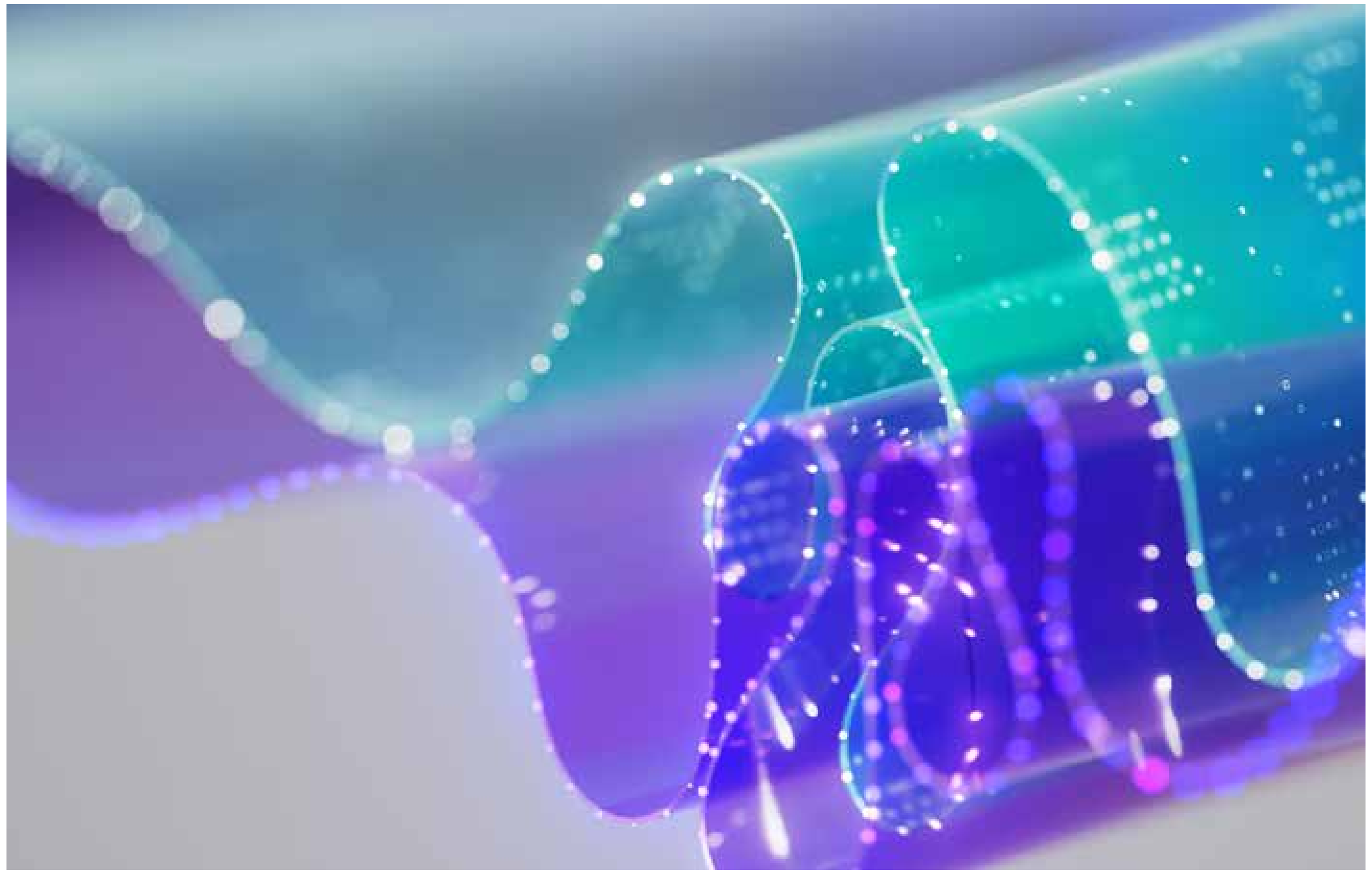
Scientific computing has developed rapidly with the emergence of computers in the last century, achieving remarkable breakthroughs in areas such as weather forecasting, oil exploration, drug design and financial analysis. It has become an essential pillar alongside theoretical and experimental research. Despite these advances, many scientific and engineering fields still grapple with challenges such as incomplete models, unclear mechanisms, and the lack of mathematical descriptions. Traditional algorithms often struggle with high-dimensional and nonlinear problems due to constraints like the curse of dimensionality. To overcome these obstacles, AI-driven machine learning has emerged as a transformative tool, offering new approaches to complex system modeling and computational optimization. By harnessing AI-driven machine learning, researchers can enhance modeling efficiency, refine workflows, and drive interdisciplinary integration.

However, the internal mechanisms and mathematical theories of AI technologies, represented by deep learning, remain underdeveloped. Many algorithms still lack rigorous proof of stability and accuracy, raising concerns about their reliability. The large-scale training of neural networks has significantly increased computational power demands, contributing to a growing tendency of 'replacing algorithm optimization with computational power'. This has led to a development model called 'brute-force', where computational power and innovation reinforce each other. High-performance computing has become a key driver in both foundational research and practical applications of AI. As AI continues to evolve, it is emerging as a crucial engine for the next generation of E-class computing.

The deep integration of AI and scientific computing raises key scientific questions, such as how to design efficient numerical algorithms to enhance the speed and accuracy of AI models; how to innovatively incorporate AI into the scientific computing tool chain; and how to harness AI to revolutionize the paradigm of numerical simulations.

When it comes to efficient numerical algorithms and model optimization, mixed-precision training and model distillation compression are driving the acceleration and deployment of AI models. Mixed-precision training dynamically integrates low- and high-precision computations based on task requirements, leveraging numerical stability and error control methods to improve computational efficiency and memory utilization. Meanwhile, model distillation uses knowledge transfer and simplifies network structures, enabling the discovery of Approximate Optimal Solutions in high-dimensional parameter spaces. By leveraging techniques such as matrix decomposition, tensor compression, and data quantization, distillation ensures that compressed models retain strong predictive accuracy, making them well suited for resource-constrained scenarios.

Innovations in scientific computing tool chains are reshaping AI's computational landscape. Current AI computations primarily rely on numerical methods such as gradient descent and matrix operations. However, these methods can lead to instability when handling ill-conditioned matrices and are difficult to interpret. Recently, researchers have explored integrating symbolic computation with numerical methods. For example, symbolic search is used to optimize matrix multiplication, while neural symbolic computing merges the advantages of data-driven learning and logical reasoning to enhance the model's interpretability and generalizability. Additionally, adaptive multi-scale computing frameworks have emerged, introducing new paradigms such as cross-scale modeling based on graph neural networks and multi-scale learning methods based on Transformers. These new technologies promise to overcome



©Just_Super / E+ / Getty

traditional limitations, steering intelligent computing toward a more stable, transparent, and generalizable future.

Traditional numerical simulation methods — such as finite differences, finite elements, and Monte Carlo — often face challenges when tackling high-dimensional and nonlinear problems, primarily due to heavy computational burdens and slow convergence rates. To address these limitations, researchers have proposed new approaches using AI tools such as deep learning and reinforcement learning. Deep neural networks are increasingly being applied in fluid mechanics, heat conduction, and structural mechanics by embedding physical laws into networks to achieve efficient approximation. Meanwhile, artificial intelligence has catalyzed the emergence of hybrid methodologies that integrate data-driven approaches with traditional simulation techniques. These fusion algorithms first perform data-informed model reduction, then synergize with classical numerical solvers, preserving physical consistency while dramatically enhancing computational efficiency. This paradigm lays the foundation for real-time monitoring and control of complex systems. Looking ahead, as AI continues to converge with physical modeling, data science, and computational mathematics, numerical simulations are set to become more efficient, intelligent, and responsive. This will lead to highly precise and efficient solutions for complex system modeling and optimization, driving advances in engineering practices and scientific discovery.

AI and scientific computing are advancing into a new phase of development, but still face challenges such as communication efficiency in distributed models, cross-scale data fusion, and stability in large-scale optimization. Solving these critical challenges will further deepen their integration, potentially overturning traditional deep learning frameworks and giving rise to more efficient, stable, and interpretable new computational methods and technologies. To fully harness the power of high-performance computing, AI algorithms must be integrated with underlying hardware architectures such as neuromorphic, optical, and quantum computing.

References

1. E, W. et al. The deep ritz method: A deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat*. **6**, 1-12 (2018).
2. Fawzi, A. et al. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature* **610**, 47-53 (2022).
3. Gao, W. et al. A mixed precision Jacobi SVD algorithm. ACM Trans. Math. Softw. 51 (2025).
4. Raissi, M. et al. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J Comput. Phys*. **378**, 686-707 (2019).

5. Complex system

Nonlinear science and complexity research encompasses a vast interdisciplinary landscape, integrating mathematics, control theory, information science, big data, and AI. Through the development of system theories, methodologies, and models, these fields seek to unravel the interactions among multiple variables, levels, and scales, shedding light on dynamic behaviours and evolutionary rules. These insights are instrumental in understanding complex systems across various domains, such as life, nature, society, engineering, economics, and management. Frontier scientific research is expanding beyond macroscopic scale, delving into super microscopic, extreme conditions, and comprehensive interdisciplinary approaches, continuously pushing the boundaries of human cognition. Complexity is the defining feature of comprehensive interdisciplinary research. Only by achieving breakthroughs in the common scientific principles of nonlinearity and complexity, these challenges can be fundamentally tackled.

With the rapid development of new-generation information technologies such as AI, big data, and supercomputing, the study of intelligent complex systems have taken on new dimensions, integrating perspectives from multiple systems and disciplines. This offers powerful new tools for addressing the challenges in nonlinear science and complexity. The 2024 Nobel Prizes in Physics and Chemistry were both awarded to scientists pioneering artificial intelligence (AI)-driven research, in recognition of their transformative breakthroughs in leveraging AI to advance fundamental studies across multiple scientific disciplines. The advancement of AI also relies on nonlinear science and complexity research to tackle data and computational power demands, as well as model interpretability, convergence, and robustness, optimizing core algorithms and architectures to drive the transformation of general AI. The development of nonlinear science and complexity research will reshape the next generation of AI development, reinforcing the symbiotic relationship in which ‘complex systems harness AI, while AI focuses on complex systems’.

Grounded in foundational mathematical theories, nonlinear science and complexity research operates across three key levels, which are understanding, mastering, and innovating. This framework fosters the development of novel research paradigms that empower each other with AI, exploring common theoretical principles and evolutionary control mechanisms across different domains. This holds strategic significance for driving technological breakthroughs in next-generation AI and addressing major challenges in human sustainable development.

The key challenges and paths to breakthroughs in this area include uncovering the common scientific principles and mathematical-physical laws that govern the evolution of various complex systems, and establishing an integrated foundational theoretical framework. To achieve breakthroughs, we should focus on evolutionary formation and dynamic structures. Specifically, we need to explore universal laws of emergent behaviour and critical states, analyse the mechanisms of self-organization and the formation of structural order, reveal the dynamic evolution laws driven by nonlinearity and randomness, study the evolutionary characteristics of high-dimensional dynamic systems, construct a structural evolution theory for multi-level complex systems, and investigate the evolution laws of information processing mechanisms and intelligent structures.

The second challenge revolves around the construction of universal models for the dynamic balance and relationship between the elements of complex systems and their structure and function, while integrating AI technology to achieve precise control of the systems. Possible breakthroughs can be explored in intelligent representation of complex systems, dynamic simulation, and optimization control. In particular, we need to research data-driven intelligent modeling and control, integrate innovative multi-modal statistical physics methods, explore multi-scale dynamic process modeling and control mechanisms, and develop phase field modeling techniques suitable for simulating complex systems.

Another hurdle is to find pathways to use complex systems theories, models, and algorithms to address AI's massive demands for data and computational power, enhance model interpretability, convergence, and robustness, while designing new core algorithms or architectures to achieve general AI. We need to explore paradigm innovation and algorithmic mechanisms for the cross-fusion of AI and complex systems. Specifically, we need to explore neural network evolution theory frameworks based on complex systems dynamics, build multi-level causal models and graph network models with structural interpretability and logical reasoning capabilities, and develop methods for heterogeneous information fusion and dynamic knowledge representation.

AI and other new intelligent technologies offer powerful tools for analysing nonlinear science and complexity challenges. Nonlinear science and complexity research are fundamental to solving comprehensive interdisciplinary scientific problems and designing new core algorithms or architectures for next-generation AI.

References

1. Stelzer, F. et al. Deep neural networks using a single neuron: folded-in-time architecture using feedback-modulated delay loops. *Nat. Commun*. **12**, 5164 (2021).
2. Floryan, D. et al. Data-driven discovery of intrinsic dynamics. *Nat. Mach. Intell*. **4**, 1113-1120 (2022).
3. Zhang, J. et al. Neural stochastic control. *NeurIPS*. **35**, 9098-9110 (2022).
4. Course, K. et al. State estimation of a physical system with unknown governing equations. *Nature* **622**, 261-267 (2023).
5. Li, X. et al. Higher-order Granger reservoir computing: simultaneously achieving scalable complex structures inference and accurate dynamics prediction. *Nat. Commun*. **15**, 2506 (2024).

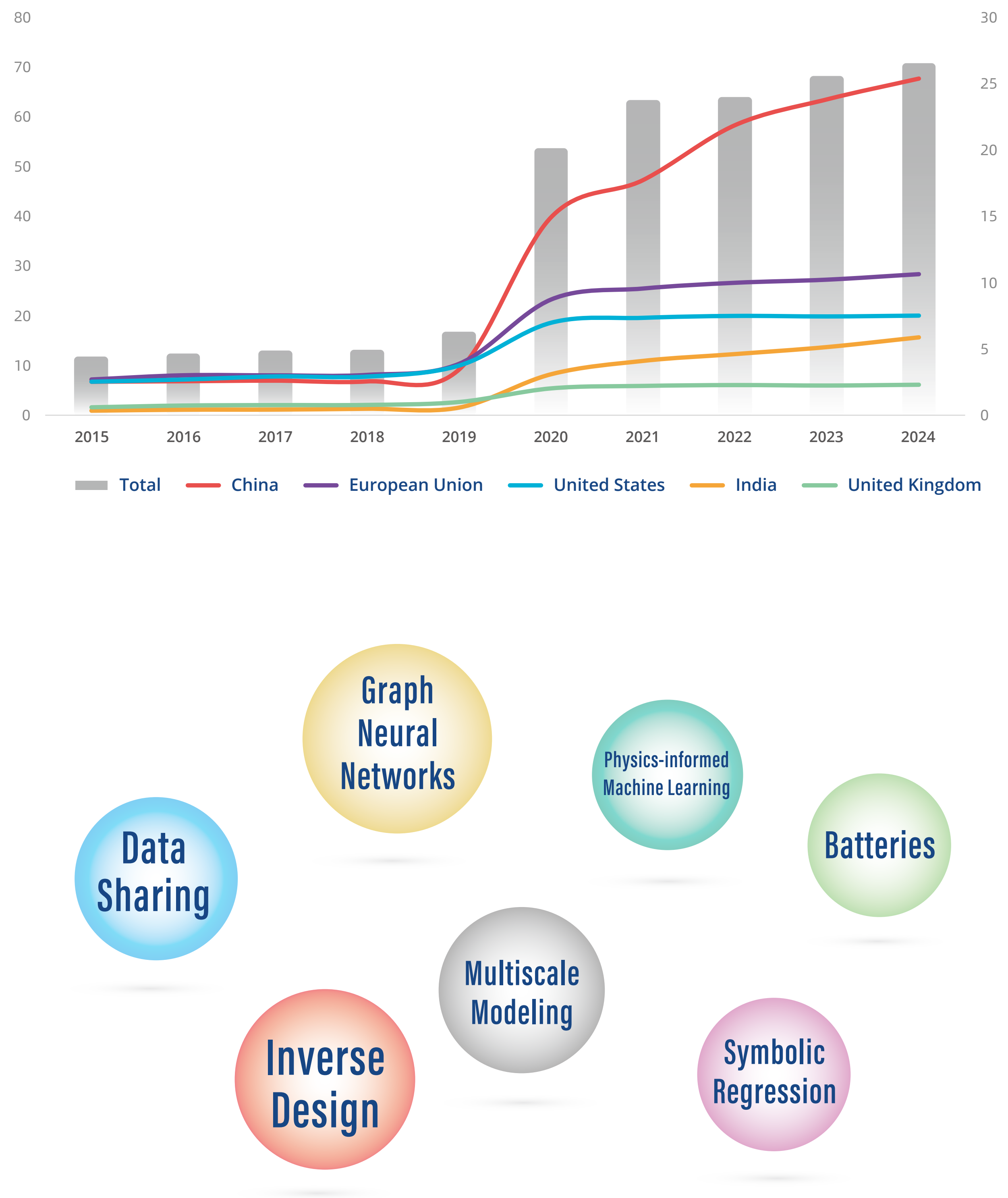
Chapter 4

AI FOR PHYSICAL SCIENCES

©Xuanyu Han / Moment / Getty

AI publications in the physical sciences began to surge in 2020, reaching a total of 70,700 by 2024 (Figure 4). In terms of national trends, China's strong ascent and India's rapid progress remain dominant patterns. Data analysis and the keyword cloud indicate that materials and battery technologies are the focal points of research. Among AI methodologies, multiscale models, symbolic regression, graph neural networks and inverse design are particularly valued by scientists, while physics-informed deep learning is driving deeper integration between AI and materials science.

FIGURE 4 | Physical Sciences – Total AI Publications, National Trends (in thousands), and Keyword Cloud (2015–2024)



1. AI for physics

1.1 Background

Physics is a foundational discipline in natural science. Its progress has long advanced through theoretical modeling, experimental verification, and computational simulations. However, with the increasing complexity of scientific questions and the exponential growth of data scales, traditional research methods face bottlenecks in both efficiency and precision. The rise of AI offers novel tools and approaches for physics, demonstrating immense potential in data-driven pattern recognition, complex system modeling, and high-efficiency algorithm optimization. The interdisciplinary field called 'AI for physics' — seeks to discover new physical laws, optimize experimental designs, accelerate material discovery, and advance theoretical innovation. Revolutionary applications of AI have already emerged in material physics, quantum mechanics, biophysics, and astrophysics.

1.2 Recent advances

1.2.1 Paradigm shifts in computational physics

Significant progress has been achieved through models like Graph Neural Networks (GNNs). For example:

- DeepMind's 'GNoME' framework predicted more than two million novel stable crystal structures[1].
- DeepMind proposed a natural excited states Variational Monte Carlo (NES-VMC) method, enabling high-precision computation of quantum excited-state wavefunctions and observables[2].
- A research team at Fudan University developed an AI model to predict electronic Hamiltonians, offering new tools for materials science[3].

1.2.2 Intelligent analysis and control in large-scale experiments

Breakthroughs have been achieved in Transformer architectures combined with contrastive learning. For example:

- A research team at Princeton University utilized AI to stabilize plasma in tokamak devices[4]; DeepMind achieved multidimensional autonomous magnetic field control of plasma in tokamaks[5].
- DeepMind introduced the AlphaQubit decoder, advancing quantum error correction[6].
- A collaborative team from the University of Toronto and UC Berkeley built an AI model to analyse radio telescope datasets for extraterrestrial signals[7].

1.2.3 Symbolic discovery of physical laws

Notable work has been done by leveraging symbolic regression and causal reasoning. For example:

- An AI model uncovered hidden physical symmetries, potentially expanding the discovery of new physics laws[8].
- A research team at MIT developed 'AI-Feynman', a physics-inspired neural network for formula recognition and theoretical derivation[9].
- A research team at Princeton University improved galaxy mass predictions by correcting astrophysical equations via symbolic regression[10].

1.3 Key challenges and paths

1.3.1 Cognitive alignment between physical priors and neural networks

Fundamental challenges include establishing differentiable inverse encoding methods to translate complex physical laws into differentiable constraints; quantifying cognitive bias to evaluate neural networks' 'depth of understanding' of physical concepts; and resolving conflicts between first principles and phenomenological models. The breakthrough strategy will involve the establishment of open physics data platforms for high-quality training datasets.

1.3.2 Emergent law mining in multi-scale systems

Key scientific challenges are cross-scale causal mapping for linking microscopic degrees of freedom to macroscopic order parameters; time-invariant neural networks for autonomous spatiotemporal scale recognition; and multimodal critical predictions for early warning near phase transitions based on microscopic fluctuations. Potential pathways to address these challenges include the development of multi-scale modeling frameworks for quantum-to-macro correlations.

1.3.3 Closed-loop systems for autonomous discovery

Critical goals may include building geometric hypothesis spaces to represent physical theories as path integrals on differentiable manifolds; deploying Bayesian experimental designs for active learning strategies to balance entropy maximization and interpretability; and generating the meta-theory for discovering fundamental physics laws. To make breakthroughs, we need to integrate physics-constrained AI models to enhance interpretability, and foster hardware-algorithm co-innovation for computational efficiency.

References

1. Merchant, A. et al. Scaling deep learning for materials discovery. *Nature* **624**, 80-85 (2023).
2. Pfau, D. et al. Accurate computation of quantum excited states with neural networks. *Science* **385**, 846 (2024).
3. Zhong, Y. et al. Accelerating the calculation of electron-phonon coupling strength with machine learning. *Nat. Comput. Sci.* **4**, 615-625 (2024).
4. Seo, J. et al. Avoiding fusion plasma tearing instability with deep reinforcement learning. *Nature* **626**, 746-751 (2024).
5. Degrave, J. et al. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* **602**, 414-419 (2022).
6. Bausch, J. et al. Learning high-accuracy error decoding for quantum processors. *Nature* **635**, 834-840 (2024).
7. Ma, P. X. et al. A deep-learning search for technosignatures from 820 nearby stars. *Nature Astronomy* **7**, 492-502 (2023).
8. He, Y.-H. AI-driven research in pure mathematics and theoretical physics. *Nat. Rev. Phy.* **6**, 546-553 (2024).
9. Udrescu, S. M. et al, AI Feynman: A physics-inspired method for symbolic regression. *Sci. Adv.* **6**, eaay2631 (2020).
10. Wadekar, D. et al. Augmenting astrophysical scaling relations with machine learning: Application to reducing the Sunyaev-Zeldovich flux-mass scatter. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2202074120 (2023).

©Tanja Ivanova / Moment / Getty

2. AI for chemistry

2.1 Background

With the advancement of data science and machine learning, AI has brought new opportunities, enabling researchers to deeply explore specialized chemistry problems while extending its influence into multiple related fields. Currently, 'AI for chemistry' is mainly applied in materials science, computational simulations, experimental process innovation, environmental science, chemical biology, and drug discovery[1].

AI technologies drive advances in chemistry through both theoretical and experimental approaches. On the theoretical side, novel AI-based algorithms significantly reduce computational resource demands by greatly enhancing simulation efficiency. Experimentally, AI facilitates laboratory automation, offering more practical and efficient solutions.

2.2 Recent advances

2.2.1 AI + Robotic chemists

AI-powered robotic chemists integrate hardware such as mobile robots and analytical instruments with automation control software, enabling broader and more flexible automated experimentation and decision-making processes in chemical synthesis. Multiple studies incorporate chemical insights provided by AI, enabling human-machine collaborative experiments, improving the efficiency and accuracy of data processing, and significantly accelerating research and development[2,3].

2.2.2 AI + Potential energy surface descriptions

Artificial neural networks based on atomistic models can construct potential energy surfaces containing energy, forces, stresses, and related information, accelerating simulation processes without sacrificing computational accuracy[4,5]. Global neural network potential methods have been successfully employed in reaction transition-state searches and reaction mechanism predictions, enhancing predictive capabilities for complicated chemical systems[6,7].

2.2.3 AI + Intelligent structural design and big data platforms

Chemical language models based on large language models have also found applications in related fields. For instance, they facilitate property prediction and structural design by converting material structures or pharmaceutical molecules into machine-interpretable information[8,9]. Additionally, algorithms such as graph neural networks and generative models can generate novel molecular or material designs from existing structures[10]. These advances have promoted the emergence of research platforms integrating big data with AI, allowing users to access knowledge databases interactively, thus supporting rapid high-throughput screening[11].

2.3 Key challenges and paths

2.3.1 Efficiency optimization of AI models in chemistry

Chemical phenomena may involve multiple scales ranging from electronic and atomic to macroscopic levels. The diversity of substances and the complexity of reactions necessitate the consumption of substantial computational and storage resources for training AI models and expanding datasets.

To achieve breakthroughs, we need to develop efficient algorithms and models tailored for simulation processes across different scales. Meanwhile, we need to design high-performance database management systems optimized for the characteristics of chemical data to enable efficient storage and rapid querying.

2.3.2 Establishment of general chemical datasets

The structural diversity of chemical substances, the complexity of experimental conditions, and the multiplicity of data sources (computational or experimental) pose significant challenges. Emphasis must be placed on data quality, ensuring fairness and completeness during data selection.

In particular, chemical data formats need to be standardized, such as representing molecular structures with SMILES and adopting the International System of Units uniformly for physical quantities. Moreover, automated tools for data cleaning, validation, and annotation from diverse sources — such as experimental results, theoretical calculations, and literature resources — need to be developed.

2.3.3 Enhancing chemistry-specific cognition in AI models

Existing large language models are primarily built on linguistic foundations. AI models for the chemical domain are required to provide enhanced scientific cognition to address more chemistry-specific problems, such as interpreting molecular and crystal structures and maintaining basic chemical literacy.

Foundational chemical scientific theories and material data need to be integrated into AI models via knowledge graphs, addressing limitations in large language models[12]. Meanwhile, the system prompt of 'solving chemistry-specific scientific problems' needs to be emphasized to ensure appropriate utilization of the relevant knowledge background.

References

1. Baum, Z. J. et al. Artificial Intelligence in Chemistry: Current Trends and Future Directions. *J. Chem. Inf. Model.* **61**, 3197–3212 (2021).
2. Dai, T. et al. Autonomous mobile robots for exploratory synthetic chemistry. *Nature* **635**, 890–897 (2024).
3. Slattery, A. et al. Automated self-optimization, intensification, and scale-up of photocatalysis in flow. *Science* **383**, eadj1817 (2024).
4. Käser, S. et al. Neural network potentials for chemistry: concepts, applications and prospects. *Digit. Discov.* **2**, 28–58 (2023).
5. Xie, X.-T. et al. LASP to the Future of Atomic Simulation: Intelligence and Automation. *Precis. Chem.* **2**, 612–627 (2024).
6. Choi, S. Prediction of transition state structures of gas-phase chemical reactions via machine learning. *Nat. Commun.* **14**, 1168 (2023).
7. Chen, D. et al. Square-pyramidal subsurface oxygen [Ag_4OAg] drives selective ethene epoxidation on silver. *Nat. Catal.* **7**, 536–545 (2024).
8. Wu, K. et al. TamGen: drug design with target-aware molecule generation through a chemical language model. *Nat. Commun.* **15**, 9360 (2024).
9. Angello, N. H. et al. Closed-loop transfer enables artificial intelligence to yield chemical knowledge. *Nature* **633**, 351–358 (2024).
10. Merchant, A. et al. Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023).
11. Ivanenkov, Y. A. et al. Chemistry42: An AI-Driven Platform for Molecular Design and Optimization. *J. Chem. Inf. Model.* **63**, 695–701 (2023).
12. Yang, L. et al. AI-assisted chemistry research: a comprehensive analysis of evolutionary paths and hotspots through knowledge graphs. *Chem. Commun.* **60**, 6977–6987 (2024).

3. AI for materials

3.1 Background

In recent years, materials science has encountered numerous bottlenecks that hinder the efficient discovery and industrialization of new materials. Traditional research methods have largely relied on empirical experiments and theoretical calculations, often requiring more than a decade to progress from conceptual validation to practical application. Furthermore, materials research is heavily dependent on expensive experimental apparatus and computational simulations, particularly in areas such as lithium-ion batteries, electrocatalysis, and advanced polymers.

Compounding these challenges is the multi-scale nature of material structures and properties, which range from electronic and atomic levels to mesoscopic and macroscopic scales. Traditional approaches struggle to construct efficient cross-scale modeling frameworks capable of capturing this complexity. The deep integration of AI with materials science holds the promise of dramatically shortening research cycles, reducing costs, and accelerating the shift toward intelligent materials science.

3.2 Recent advances

3.2.1 Materials design

The field of materials design is undergoing a transformative shift, moving from traditional high-throughput screening to machine learning-driven inverse design. Generative models have significantly expanded the boundaries of traditional materials search space, establishing themselves as a key technology in the discovery of novel materials.

Key methodologies driving this innovation include variational autoencoders, generative adversarial networks, diffusion models, and large language models (LLMs). For example, MatterGen[1] employs diffusion models to generate stable materials.

3.2.2 Material property prediction

Advanced methods for predicting material properties increasingly leverage graph neural networks, multi-fidelity learning and deep learning Hamiltonian approaches. These techniques offer a balance between high accuracy and the scalability required for large-scale material screening.

In cutting-edge fields such as batteries, catalysis, polymers and optoelectronic materials, these methods are being deeply integrated to acceleration identification of high-value materials.

3.2.3 AI agents and large models in materials science

AI agents and large models are transforming materials science by leveraging advanced technologies such as graph neural networks and LLMs to enable precise predictions of key performance indicators and ensure reliable scientific reasoning.

For instance, the MatChat AI agent[2], built on an extensive corpus of academic literature, facilitates traceable Q&A in materials science. Similarly, the MatterChat large model[3] bridges high-resolution atomic structural data with textual representations from LLMs, significantly enhancing predictive accuracy while offering a more intuitive human–machine interaction interface.

3.2.4 Autonomous laboratories for intelligent material synthesis

Intelligent synthesis in materials science is evolving at an unprecedented speed, fueled by large-scale DFT databases, generative structure prediction techniques and LLMs. The integration of robotic automation with cloud-based scheduling optimizes experimental workflows and ensures real-time model updates.

For example, the autonomous laboratory A-Lab[4] employs a DFT database and natural language models to drive closed-loop inorganic powder synthesis. Meanwhile, cloud-based asynchronous distributed collaboration enables experimental platforms to operate independently under centralized AI scheduling[5].

3.3 Key challenges and paths

3.3.1 Materials database and data sharing

Reliable, and multi-source heterogeneous materials databases need to be established, and global data sharing should be facilitated.

The heterogeneity of data sources in materials science — coupled with inconsistent data quality and limited sharing practices — constrains the generalization capabilities of AI models and undermines the credibility of material predictions.

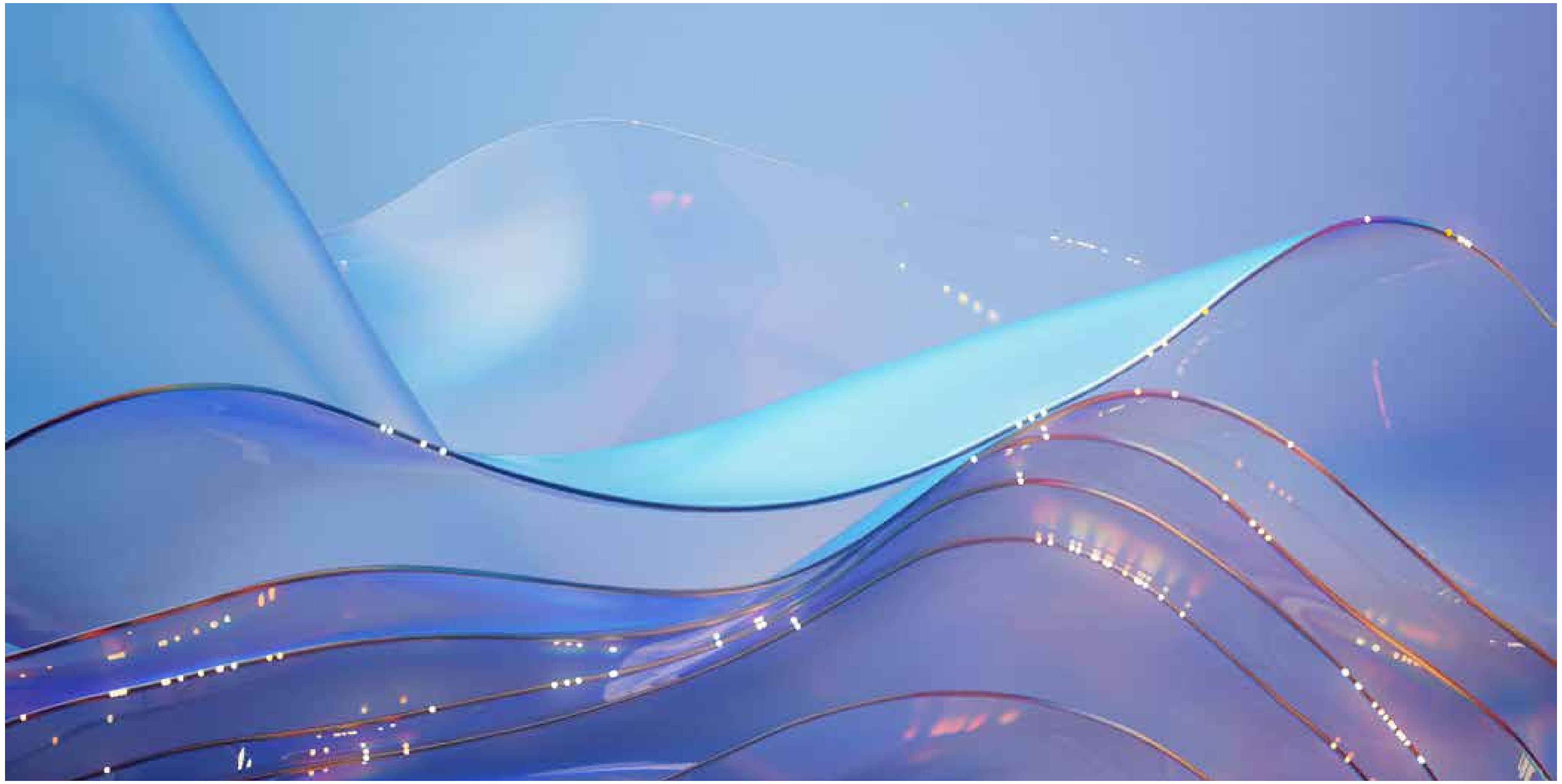
To solve this, data storage formats need to be standardized to ensure compatibility across computational, experimental, and AI simulation platforms. Multi-fidelity datasets also need to be constructed by cleansing, completing, and deduplicating data to enhance overall quality. Furthermore, an open ecosystem for materials big data needs to be established to facilitate high-quality global data sharing.

3.3.2 Feasible and scalable generative AI materials models

Chemically feasible and scalable generative AI materials models need to be developed for multi-scale autonomous design.

While generative AI has shown promise in accelerating materials discovery, current approaches often overlook critical factors such as thermodynamic stability, synthesis pathways, and device-level performance correlations. This gap presents challenges in achieving seamless integration from atomic-scale materials design to macroscopic system optimization.

Physical constraints need to be integrated into generative models, and screening mechanisms such as DFT need to incorporated to improve predictions of thermodynamic stability. Large language



©shulz / E+ / Getty

models also need to be combined with autonomous laboratory workflows to optimize synthesis pathways and assess experimental feasibility.

3.3.3 AI and autonomous laboratories for automated materials discovery

AI needs to be integrated with autonomous laboratories to achieve fully automated material discovery and intelligent performance optimization.

The integration of AI-driven computations, experimental automation, and multi-scale modeling can accelerate the design of groundbreaking materials, such as room-temperature superconductors, self-healing flexible semiconductors, and ultra-light, high-strength nanocomposites.

To achieve this, generative models can be used to propose candidate materials that meet target performance criteria, followed by precise screening processes. Robotic synthesis and intelligent characterization for fully automated material fabrication can be leveraged, and reinforcement learning and adaptive optimization should be employed to continuously refine materials design.

In recent years, materials science has encountered numerous bottlenecks that hinder the efficient discovery and industrialization of new materials. Traditional research methods have largely relied on empirical experiments and theoretical calculations, often requiring more than a decade to progress from conceptual validation to practical application. Furthermore, materials research is heavily dependent on expensive experimental apparatus and computational simulations, particularly in areas such as lithium-ion batteries, electrocatalysis, and advanced polymers.

Compounding these challenges is the multi-scale nature of material structures and properties, which range from electronic and atomic levels to mesoscopic and macroscopic scales. Traditional approaches struggle to construct efficient cross-scale modeling frameworks capable of capturing this complexity. The deep integration of AI with materials science holds the promise of dramatically shortening research cycles, reducing costs, and accelerating the shift toward intelligent materials science.

References

1. Zeni, C. et al. A generative model for inorganic materials design. *Nature* 639, **624**-632 (2025).
2. Chen, Z. Y. et al. MatChat: A large language model and application service platform for materials science. *Chinese Physics B* **32**, 118104 (2023).
3. Tang, Y. et al. MatterChat: A Multi-Modal LLM for Material Science. *arXiv preprint* **arXiv:2502.13107** (2025).
4. Szymanski, N. J. et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**, 86-91 (2023).
5. Strieth-Kalthoff, F. et al. Delocalized, asynchronous, closed-loop discovery of organic laser emitters. *Science* **384**, eadk9227 (2024).

4. AI for energy

4.1 Background

As global energy demand continues to rise and environmental pressures intensify, traditional research paradigms face mounting challenges in improving efficiency and driving innovation in energy materials. Despite significant efforts, progress in energy material research remains slow, approaching theoretical performance limits.

For example, in catalytic materials discovery, even with robot-assisted experimental synthesis that compresses synthesis, characterization, and testing cycles to just one week, it would still take more than a century to screen 5,000 potential material combinations[1]. This highlights the limitations of existing methodologies in meeting the urgent need for low-carbon, cost-effective, and high-performance energy materials — particularly in light of dual carbon strategic objectives.

4.2 Recent advances

Breakthrough innovations in novel energy material design are needed, and AI-driven approaches are emerging as a transformative solution. By accelerating molecular discovery, optimizing material design, and significantly improving R&D efficiency[2], AI-empowered methodologies hold promise to help the world achieve carbon peaking by 2030 and the subsequent carbon neutrality targets.

Machine learning is revolutionizing energy materials research, particularly in energy storage and catalysis, transforming material development from trial-and-error approaches to a closed-loop ‘design-verification’ paradigm. This shift is driven by accelerated big data screening, deeper exploration of complex structure-activity relationships, and intelligent optimization of material parameters.

Recent breakthroughs highlight the impact of machine learning across various domains:

▪ Yang Cao at the University of Toronto, Ontario, Canada, and his team[3] developed a generative reinforcement learning framework to identify novel organic candidates for redox flow batteries by optimizing molecular stability and redox potentials.

▪ Huisheng Peng at Fudan University, Shanghai, China and the research team[4] integrated unsupervised learning with cheminformatics to extract critical molecular fragments from electrochemical features, constructing a high-quality battery database and successfully designing lithium-ion carrier molecules.

▪ Guangmin Zhou at Tsinghua University, Beijing, China and the research team[5] combined unsupervised learning with nudged elastic band simulations to enable rapid screening of lithium-ion conductors in solid-state batteries, establishing a predictive model for high-conductivity electrolytes based on limited computational data.

▪ Jianchang Wu at Helmholtz-Institute Erlangen–Nürnberg (HI-ERN), Erlangen, Germany, and the team[6] utilized Bayesian optimization to align high-throughput synthesis data of organic semiconductors, developing superior hole-transport materials for solar cells.

▪ Jinlan Wang at Southeast University, Nanjing, China and the team[7] proposed a machine learning-assisted synthesis framework for 2D perovskites, combining experimental data with chemical prior knowledge to significantly enhance new material synthesis efficiency.

▪ Nicholas Jackson at University of Illinois at Urbana-Champaign, Urbana, the United States, and the team[8] innovatively introduced a closed-loop transfer research method that iteratively optimizes molecular photostability through Bayesian optimization coupled with machine learning hypothesis validation.

▪ Yujie Xiong at University of Science and Technology of China, Hefei, China and the team[1] established a machine learning-driven high-throughput screening protocol which combines photosensitization, electron transfer, and catalytic descriptors to optimize efficient molecular photocatalysts for CO_2reduction.

- Sargent Edward at University of Toronto, Ontario, Canada, and the team[9] constructed an ML-DFT collaborative feedback framework to design low-energy-barrier catalytic structures by machine-learning surface electronegativity and CO_2 adsorption energy.
- Shizhang Qiao at the University of Adelaide, Australia, and the team[10] integrated 2D-3D machine learning algorithms to create a dataset covering all C-C coupling precursors and active sites, enabling rapid mechanistic analysis of complex electrocatalytic coupling reactions.

4.3 Key challenges and paths

The key challenges, however, include:

4.3.1 Data silos hinder R&D efficiency.
The dispersed and non-standardized multi-source heterogeneous data — including experimental results, physical properties, and behaviour characteristics — for energy materials poses a significant challenge to cross-domain data integration, impeding the construction of large-scale databases and knowledge transfer. It's crucial to establish a standardized evaluation system linking 'microstructure-macro-performance'. This system should integrate physics-based models and data-driven approaches to extract transferable structural descriptors. Furthermore, establishing a multi-dimensional validation platform that encompasses key parameters — such as electrochemical properties and catalytic activity — will enable a closed-loop R&D cycle of computational prediction, experimental validation, and model iteration.

4.3.2 Model opacity limits mechanistic understanding.
The disconnect between deep learning predictions and underlying microscopic mechanisms makes it difficult to interpret structure-property relationships, limiting the theoretical insights needed to guide new material design. To address this issue, integrating fundamental physicochemical constraints within model architectures is essential, which will pave the way for the development of data-mechanism driven molecular generative models.

4.3.3 Multi-objective optimization is often trapped in local optima.
Current AI algorithms often struggle with multi-objective optimization, frequently converging to local optima when balancing competing factors such as conductivity, stability, and cost. This limitation restricts their applicability in complex scenarios. It's essential to integrate key macroscopic indicators and microscopic features associated with energy conversion, storage, and utilization scenarios to construct a cross-media, multi-scale energy molecule database. Developing physics-constrained generative models will enable Pareto frontier exploration and global optimization within the chemical space.

4.3.4 Integration of experimental and design loops is ineffective.
Feedback delays between AI predictions and experimental validation, coupled with insufficient synergy between high-throughput computational and automated experimental platforms, hinders iterative efficiency. It's crucial to move beyond traditional linear R&D pathways and establish a cross-dimensional framework that encompasses hypothesis generation, intelligent screening, and targeted synthesis. Meanwhile, transfer learning can facilitate knowledge reuse in small-data scenarios.

4.3.5 The safety assessment framework is absent.
The lack of quantitative standards for safety indicators, such as environmental toxicity and long-term stability, and insufficient sustainability constraints within AI optimization frameworks, raises a significant concern. It's essential to integrate proactive safety assessments into the AI design process. By incorporating life cycle analysis and materials genome engineering, researchers can develop multi-objective optimization algorithms that balance performance and sustainability.

References

1. Hu, Y. et al. Identifying a highly efficient molecular photocatalytic CO2 reduction system via descriptor-based high-throughput screening. *Nat. Catal.* **8**, 126–136 (2025).
2. Yao, Z. et al. Machine learning for a sustainable energy future. *Nat. Rev. Mater.* **8**, 202–215 (2023).
3. Cao, Y. et al. Reinforcement learning supercharges redox flow batteries. *Nat. Mach. Intell.* **4**, 667–668 (2022).
4. Chen, S. et al. External Li supply reshapes Li deficiency and lifetime limit of batteries. *Nature* **638**, 676–683 (2025).
5. Lao, Z. et al. Data-driven exploration of weak coordination microenvironment in solid-state electrolyte for safe and energy-dense batteries. *Nat. Commun.* **16**, 1075 (2025).
6. Wu, J. et al. Inverse design workflow discovers hole-transport materials tailored for perovskite solar cells. *Science* **386**, 1256–1264 (2024).
7. Wu, Y. et al. Universal machine learning aided synthesis approach of two-dimensional perovskites in a typical laboratory. *Nat. Commun.* **15**, 138 (2024).
8. Angello, N. H. et al. Closed-loop transfer enables artificial intelligence to yield chemical knowledge. *Nature* **633**, 351–358 (2024).
9. Zhong, M. et al. Accelerated discovery of CO2 electrocatalysts using active machine learning. *Nature* **581**, 178–183 (2020).
10. Li, H. et al. Machine Learning Big Data Set Analysis Reveals C–C Electro-Coupling Mechanism. *J. Am. Chem. Soc.* **146**, 22850–22858 (2024).

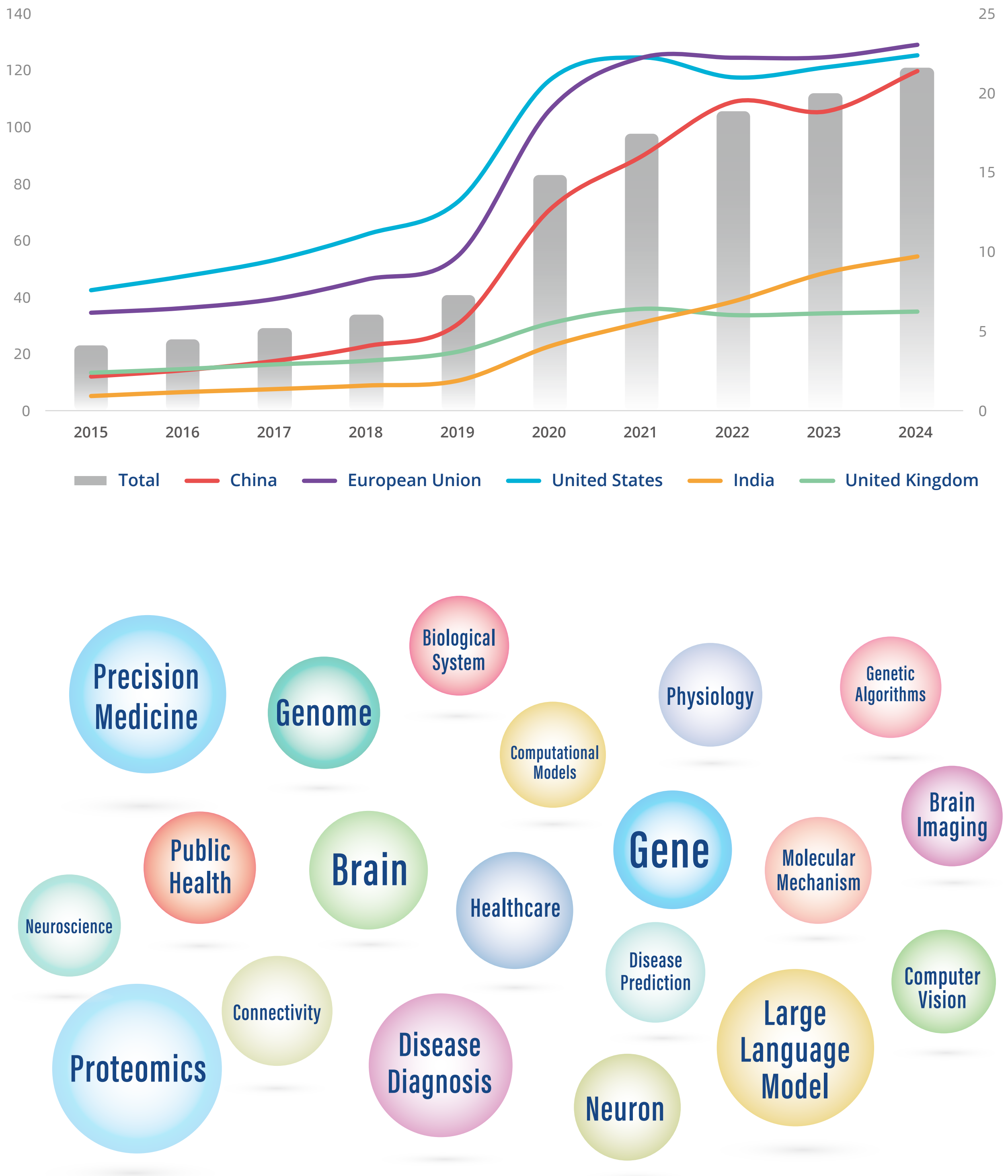
Chapter 5

AI FOR LIFE SCIENCES

©Andrew Brookes / Connect Images / Getty

The rapid growth of AI publications in the life sciences also began around 2020, with global output soaring to 120,700 by 2024 (Figure 5). Although the US and EU have long dominated research in this field, China has rapidly closed the gap, with its 2024 publication volume nearly on par with both. Data analysis and the keyword cloud underscore neuroscience, genomics and health as the most prominent areas of research. AI is accelerating drug discovery through large language models, advancing disease diagnosis with high-resolution imaging, and driving deep integration of multiscale data-driven models with genomics and proteomics — collectively advancing our understanding of complex biological systems.

FIGURE 5 | Life Sciences – Total AI Publications, National Trends (in thousands), and Keyword Cloud (2015–2024)



1. AI for synthetic biology

1.1 Background

Synthetic biology is an interdisciplinary field that combines biology, engineering and computational science, to design and create novel biological systems with specific functions. Recent advances in gene editing, de novo protein design, and metabolic engineering have accelerated synthetic biology, leading to transformative applications in healthcare, environmental remediation and advanced materials. Artificial intelligence (AI), with its robust multitasking learning capabilities and capacity for intelligent exploration of uncharted spaces, is becoming increasingly integral to synthetic biology. AI facilitates the systematic decoding of the intricate relationships among biological sequences, structures and functions, enhancing efficiency and precision in design. This synergy is shifting the field from an era of biological modification to a new phase of biological creation.

1.2 Recent advances

1.2.1 AI-powered gene editing and nucleic acid vaccines

Through deep learning and large-scale genomic data analysis, AI can identify therapeutic targets within complex genetic datasets and accurately predict the biological effects of gene editing and antigen design. This significantly enhances the precision and efficiency of molecular regulation, and programming of cellular functions[1]. AI-driven optimization of CRISPR-based editing strategies and mRNA-based vaccine antigen design is offering innovative solutions for precision medicine.

1.2.2 AI-driven de novo protein

Proteins are the central executors of biological functions and play a pivotal role in synthetic biology. AlphaFold has revolutionized protein structure prediction, achieving atomic-level accuracy and covering 98.5% of the human proteome[2]. Diffusion model-based tools such as RFdiffusion provides a novel approach to de novo protein design[3]. Meanwhile, natural language model-based tools such as ProGen[4] and EVOLVEpro[5] enable the de novo synthesis and directed evolution of functional proteins, including enzymes and therapeutic antibodies.

1.2.3 AI-transformed bio-manufacturing

Through AI-driven data analysis and model optimization, researchers can precisely design and optimize synthetic pathways for target compounds, enabling tailored engineering of microbial genomes to achieve high-yield production of bioproducts such as artemisinin, terpenoids and their derivatives. Tools such as Evo mark a paradigm shift from localized pathway optimization to whole-genome design, enabling the generation of entirely novel DNA sequences with tailored functions and even the construction of complete microbial genomes[6]. These innovations pave the way for efficient, customizable microbial factories with broad industrial applications.

1.3 Key challenges and paths

1.3.1 Design and scalability of complex biological systems

Despite progress in the de novo design of individual biomolecules (genes, proteins, lipids), the design and assembly of complex biological systems (nanorobots, eukaryotic cells) remain challenging due to limited scalability, hindering large-scale production.

The paths to breakthroughs include:

- Develop AI-driven modular design tools to automate the assembly and optimization of complex biological systems.
- Establish evolutionary algorithm-based adaptive optimization models to enhance system iterability and scalability.

1.3.2 Customized design and application of therapeutic functional proteins

Although AI is capable of designing novel functional proteins, their therapeutic application is constrained by immune recognition and clearance. AI prediction accuracy is also inadequate for protein-protein interactions.

The paths to breakthroughs include:

- Improve AI algorithms by developing multimodal deep learning models to achieve high-precision prediction of interaction interfaces and functional design.
- Engineer 'immune-stealth' proteins that dynamically shield immunogenic epitopes via environmental responsiveness, ensuring long-term stability *in vivo*.

1.3.3 Biomolecular integrated circuits and bio-computing machines

The limited integration density and autonomous operation capability of current biocomputing hardware restrict its potential in high-parallel computing and complex system optimization.

The paths to breakthroughs include:

- Develop AI-based modular design tools and integration systems for biomolecular components to enhance circuit scale and functionality.
- Advance bidirectional electronic-molecular communication technologies to improve computational bandwidth and efficiency, enabling fully autonomous bio-computing machines.

References

1. Gosai, S. J. et al. Machine-guided design of cell-type-targeting cis-regulatory elements. *Nature* **634**, 1211-1220 (2024).
2. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596 (2021).
3. Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089-1100 (2023).
4. Madani, A. et al. Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **41**, 1099-1106 (2023).
5. Jiang, K. et al. Rapid in silico directed evolution by a protein language model with EVOLVEpro. *Science* **387**, eadr6006 (2025).
6. Nguyen, E. et al. Sequence modeling and design from molecular to genome scale with Evo. *Science* **386**, eado9336 (2024).

2. AI for medicine

2.1 Background

AI has evolved significantly in medicine, transitioning from early rule-based systems to deep learning frameworks. Its applications have extended from information technology to personalized diagnosis and treatment[1]. Early AI applications were focused on computer-aided diagnosis and clinical decision support systems, improving the accuracy of diagnosis and treatment through image recognition and data analysis. A major breakthrough in 2017 introduced Transformer architecture, which enabled the development of pre-trained models such as BERT and GPT. These models significantly improved the ability to process and interpret medical texts, laying the groundwork for Large Language Models (LLMs) and Multimodal Large Language Models (MLLMs).

Currently, LLMs can process complex medical texts, including medical record interpretation, literature analysis, and reasoning-based clinical decisions[2]; MLLMs further integrate text, image, genome and other multi-source data to build a cross-modal diagnosis and therapeutic framework to improve diagnostic efficiency[3].These technological breakthroughs are particularly valuable in resource-limited areas, where automated analysis can reduce healthcare costs and alleviate expertise shortages. Meanwhile, collaborative innovation in the global open-source community has accelerated the multilingual adaptation and international application of large-scale medical models, unleashing the potential for clinical transformation[4].

2.2 Recent advances

2.2.1 Medical LLMs

LLMs have demonstrated strong capabilities in parsing complex medical texts, enabling extraction of key information from electronic medical records, PubMed documents and guidelines. This supports the matching of clinical trials and construction of knowledge graphs. In terms of interactive applications, LLM-driven chatbots (e.g., Woebot) have shown efficacy in delivering cognitive behavioral therapy for mental health interventions[5]. Between 2023 and 2024, major technology companies have launched medical-specific LLMs. Google's Med-PaLM2 has passed the United States Medical Licensing Examination, reaching an expert-level performance[6]; Microsoft's BioGPT focuses on biomedical text generation, and its F1-score performance in relationship extraction tasks is outstanding[7]. LLMs have demonstrated unique advantages in integrating genetic variation data to improve the accuracy and generalizability of gene expression prediction, providing new paths for elucidating the genetic mechanisms of complex diseases and identifying novel drug targets, thereby accelerating drug development.

2.2.2 Medical MLLMs

MLLMs facilitate multidimensional analysis of disease mechanisms by fusing imaging, genetic and clinical data to empower precision medicine. For example, the MuMo model, developed by Peking University, utilized multimodal data — including imaging, pathological and clinical information — from 429 HER2-positive gastric cancer patients to guide therapy and predict immunotherapy response[8]. MedFound-DX-PA achieved an 80.7% disease diagnosis accuracy rate under zero-shot learning[9]. MLLMs have gradually penetrated clinical practice, covering the entire process from disease screening, diagnosis to personalized management.

2.3 Key challenges and paths

2.3.1 The knowledge system of medical models lacks breadth and depth of integration

Knowledge needs to be integrated to cope with the complexity of medicine: the breadth requires the integration of cross-domain data such as electronic medical records, imaging and genomes, and the depth requires expert understanding of disease mechanisms and individualized diagnosis and treatment. Data privacy constraints, regional standards variability and delays in knowledge updates limit global applicability. A dynamic knowledge system should be established to ensure comprehensiveness and accuracy.

The paths to breakthroughs include:

- Employ Natural Language Processing and knowledge graph to build a dynamic system[10].
- Utilize multimodal fusion to improve diagnostic accuracy.
- Implement distributed learning to achieve global sharing under privacy protection and establish a medical knowledge hub.

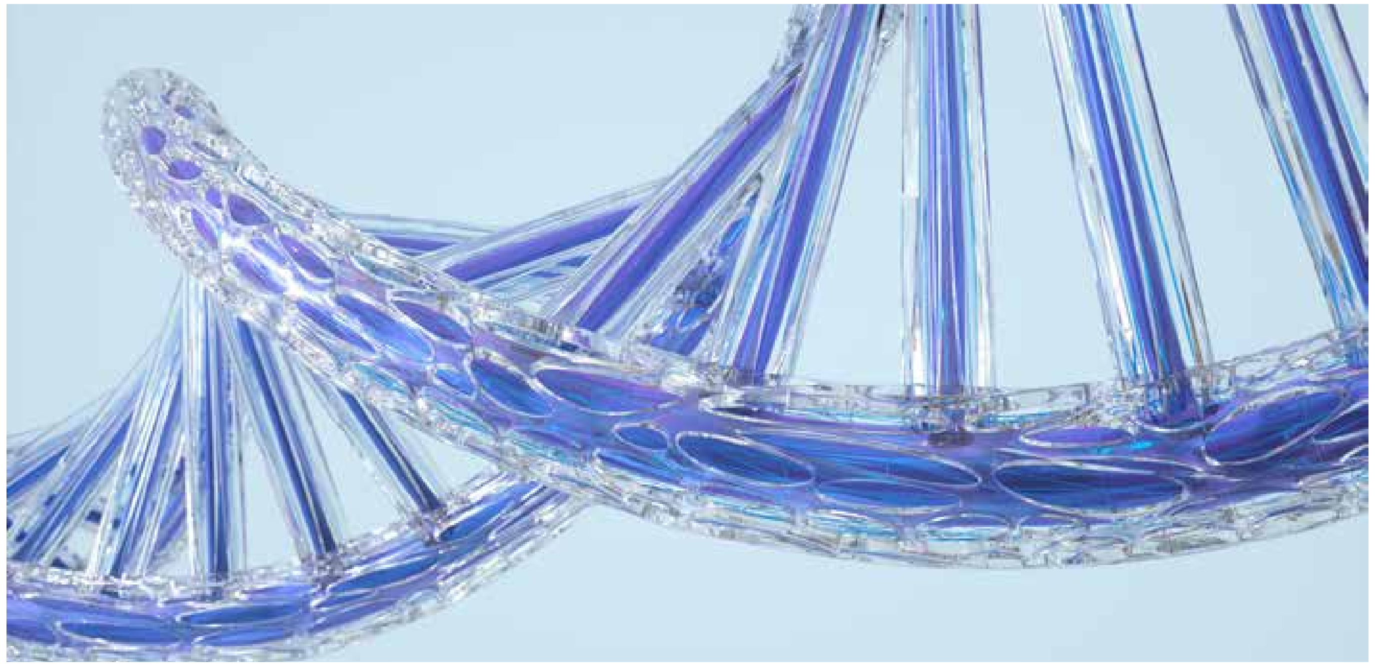
2.3.2 Lack of evidence chains and reasoning in clinical decision-making models

Many existing models fail to provide a complete chain of evidence, and the conclusions often lack transparent reasoning, which affects trust in doctors, especially in high-risk scenarios. Multimodal data fusion (genomics, clinical records, imaging, lifestyle) requires stronger reasoning capabilities. An interdisciplinary reasoning mechanism should be built to deal with complex diagnosis and treatment.

The paths to breakthroughs include strengthening causal reasoning to extract key evidence11; building a multi-agent system to simulate multidisciplinary consultation; and combining reliable data sources to generate a visual evidence chain to meet regulatory and clinical needs.

2.3.3 Model interpretability and reliability

The black-box nature of AI systems reduces clinical trust[12]. The risk of generating wrong information remains a concern. While regulators require transparent decision-making, there is often a trade-off between model explainability and accuracy. It is necessary to establish a verifiable reasoning process and improve transparency through visualization.



©Andriy Onufriyenko / Moment / Getty

The paths to breakthroughs include:

- Integrate knowledge graphs and visualization techniques such as heatmap positioning to explain decision logic.
- Develop interactive interfaces to support parameter adjustment.
- Draw on approaches such as the DR.KNOWS model to improve diagnostic traceability[13].

2.3.4 Ethical and privacy risks

Training data often involves sensitive patient information, such as images and genomic data, raising privacy concerns. Furthermore, it is necessary to clarify the responsibility for AI misdiagnosis, address technology fairness in resource-poor areas, and avoid exacerbating medical inequality[4]. Clinical deployment needs to balance informed consent, physician training, and fairness.

The paths to breakthroughs include:

- Employ federated learning[4], differential privacy and homomorphic encryption to build a protection system.
- Formulate global ethics guidelines for AI in medicine, give priority to localized deployments that comply with Health Insurance Portability and Accountability Act[14].
- Promote trust through comprehensive ethics training and governance.

References

1. Kaul, V. et al. History of artificial intelligence in medicine. *Gastrointest. Endosc.* **92**, 807–812 (2020).
2. Wang, D. et al. Large language models in medical and healthcare fields. *Artif. Intell. Rev.* **57**, 299 (2024).
3. Qiu, J. et al. The application of multimodal large language models in medicine. *Lancet Reg. Health West. Pac.* **45**, 101048 (2024).
4. Qiu, P. et al. Towards building multilingual language model for medicine. *Nat. Commun.* **15**, 8384 (2024).
5. Fitzpatrick, K. K. et al. Delivering Cognitive Behavior Therapy Using Woebot. *JMIR Ment. Health* **4**, e19 (2017).
6. Singhal, K. et al. Toward expert-level medical question answering with large language models. *Nat. Med.* **31**, 943–950 (2025).
7. Luo, R. et al. BioGPT: generative pre-trained transformer for biomedical text generation. *Brief. Bioinform.* **23**, bbac409 (2022).
8. Chen, Z. et al. Predicting gastric cancer response to anti-HER2 therapy. Signal Transduct. *Target. Ther.* **9**, 222 (2024).
9. Liu, X. et al. A generalist medical language model for disease diagnosis assistance. *Nat. Med.* **31**, 932–942 (2025).
10. Christophe, C. et al. Med42--evaluating fine-tuning strategies. *arXiv preprint* **arXiv:2404.14779** (2024).
11. Jiang, P. et al. Reasoning-Enhanced Healthcare Predictions. *arXiv preprint* **arXiv:2410.04585** (2024).
12. Li, J. et al. Integrated image-based deep learning for diabetes care. *Nat. Med.* **30**, 2886–2896 (2024).
13. Gao, Y. et al. Leveraging medical knowledge graphs into large language models for diagnosis prediction: Design and application study. *arXiv preprint* **arXiv:2308.14321** (2025).
14. Mehandru, N. et al. Evaluating large language models as agents in the clinic. *npj Digit. Med.* **7**, 84 (2024).

3. AI for neuroscience

3.1 Background

The deep integration of AI and neuroscience is advancing our understanding of brain mechanisms at an unprecedented pace. Neuroscience, by analysing the brain's structure, functions and cognitive patterns, has provided insights that fueled the development of algorithms such as perceptual networks and spiking neural networks[1]. At the same time, AI is accelerating the development of brain imaging, connectomics, and neural computational models, enhancing the diagnosis and treatment of neurological disorders[2].

In recent years, techniques such as automated image segmentation, 3D reconstruction and multimodal data fusion have enabled researchers to reconstruct complex neural networks from C. elegans to human cerebral cortex samples, demonstrating AI's tremendous potential in handling vast brain data[3,4]. This bidirectional empowerment drives breakthroughs in fundamental science, and lays a solid foundation for clinical applications and the development of brain-inspired intelligence[5].

3.2 Recent advances

AI technologies have recently made significant progress in various areas of neuroscience, particularly in data acquisition, processing and analysis. In the field of connectomics, high-resolution imaging combined with automated image processing has enabled researchers to reconstruct neural networks from small organisms to human cerebral cortex samples. For example, the collaboration between Harvard University and Google[6], which used serial-section electron microscopy and automated image segmentation to reconstruct a 3D model of a 1mm^3 human cortical sample, recording approximately 150 million synapses[2,3]. The US BRAIN Initiative, produced a panoramic neuronal atlas using multimodal optical microscopy, spanning submicron resolution to whole-brain scale[6].

AI has been widely used for the automatic segmentation and anomaly detection of multimodal imaging such as MRI, PET and EEG, significantly improving the early diagnosis accuracy of neurodegenerative conditions like Alzheimer's disease[1,7]. In neural signal decoding and brain-computer interface (BCI) technologies, deep learning algorithms have enabled real-time decoding of complex motor intentions, such as imagined handwriting and fine finger movements in people with paralysis[8,9].

Furthermore, multimodal data fusion and cross-scale modeling offer new perspectives for integrating anatomical, functional and molecular information. Recently, Jiang et al. introduced the NeuroXiv platform, enabling dynamic mining of whole-brain neuromorphological data[10]; Liu et al. revealed the intrinsic relationship between neuronal diversity and network stability using whole-brain morphological measurements, providing a theoretical foundation for biologically plausible neural network model design[6]. Meanwhile, the ARNI Institute, in the UK, developed a connectome-driven neural architecture search method, opening new directions for brain-inspired computing[11].

3.3 Key challenges and paths

3.3.1 Intelligent data generation and annotation

Advanced imaging techniques such as electron microscopy, optical microscopy, and diffusion MRI generate vast and multimodal datasets. A major challenge is how to leverage AI techniques—such as generative adversarial networks (GANs)—to automate data generation and intelligent annotation, accurately capturing the microstructural and functional features of neural networks.

The paths to breakthroughs include:

- Develop AI-driven neuroinformatics platforms to integrate real-time multimodal data streams and seamlessly fuse various data types.
- Design cross-modal data generation frameworks based on GANs to enable conversion and supplementation between different imaging modalities.
- Create semi-automated labeling systems combining contrastive learning and active learning to reduce manual annotation workload.

3.3.2 Novel brain-computer interfaces and high-precision neural signal decoding

Brain–computer interface (BCI) technologies are evolving toward non-invasive and minimally invasive paradigms. A critical challenge is achieving real-time, high-precision decoding of complex neural signals through multimodal signal integration, advancing human-computer interaction technologies.

The paths to breakthroughs include:

- Integrate multimodal data to develop adaptive decoding algorithms that account for the dynamic, non-stationary nature of neural activity.
- Build high-level cognitive state decoding frameworks to enable cross-level semantic decoding—from low-level motor intentions to higher-order abstract thoughts (memory)—paving the way for brain-to-brain communication and AI-augmented memory enhancement technologies.

3.3.3 Building interpretable and biologically-inspired AI models

To move beyond the black box limitation of current AI models, it is important to design models that are both predictive and interpretable while biologically grounded. Such models could not only help reveal the decision-making mechanisms underlying cognition, but offer more intuitive frameworks for clinical diagnosis and treatment.

The paths to breakthroughs include:

- Develop hybrid architectures that integrate symbolic AI, graph neural networks, and traditional deep learning to map AI predictions to understandable neurobiological concepts and uncover underlying biological mechanisms.
- Construct AI-driven models capable of predicting individual responses to neural interventions, allowing for efficient adaptation from large datasets to personalized cases without extensive retraining.

3.3.4 Multiscale, multimodal brain structure and functional network models with cross-level data integration

Information processing spans from single neurons to whole-brain networks. Building comprehensive models that encompass cellular, regional, and whole-brain levels is key to understanding higher-order cognitive functions and behaviours.

The paths to breakthroughs include:

- Construct multi-layer, multiscale integrated brain atlases that synchronize structural and functional data across different modalities on a unified temporal axis, overcoming current limitations in data fusion.
- Develop hierarchical dynamic brain network models (cell-circuit-system) to enable bidirectional information flow across different brain scales.

Research addressing these frontier challenges will provide both conceptual frameworks and practical tools for the next-generation brain-inspired intelligence systems and precision medical technologies.

By accelerating brain data acquisition, integration and intelligent analysis, AI deepens understanding of the nervous system's fundamental principles, and drives progress in neurodegenerative disease research, brain-computer interfaces and brain-inspired computing.

References

1. Onciul, R. et al. Artificial intelligence and neuroscience: transformative synergies in brain research and clinical applications. *J. Clin. Med*. **14**, 550 (2025).
2. Manning, A. J. Epic science inside a cubic millimeter of brain-Researchers publish largest-ever dataset of neural connections. *Harvard Gazette*. (2024).
3. Max Planck Institute for Brain Research. Faster reconstruction of the connectome. *Press Release*. (2015).
4. Park, C. et al. Automated synapse detection method for cerebellar connectomics. *Front. Neuroanat*. **16**, 760279 (2022).
5. Fuller-Wright, L. Mapping an entire (fly) brain: A step toward understanding diseases of the human brain. *Princeton Univ. News*. (2024).
6. Liu, Y. et al. Neuronal diversity and stereotypy at multiple scales through whole brain morphometry. *Nat. Commun*. **15**, 10269 (2024).
7. Qiu, S. et al. Multimodal deep learning for Alzheimer's disease dementia assessment. *Nat. Commun*. **13**, 3404 (2022).
8. Willett, F. R. et al. High-performance brain-to-text communication via handwriting. *Nature*, **593**, 249-254 (2021).
9. Willsey, M. S. et al. A high-performance brain-computer interface for finger decoding and quadcopter game control in an individual with paralysis. *Nat. Med*. **31**, 96-104 (2025).
10. Jiang, S. et al. NeuroXiv: AI-powered open databasing and dynamic mining of brain-wide neuron morphometry. *Nat. Methods* DOI: 10.1038/s41592-025-02687-2. (2025).
11. ARNI Institute. Connectome-Guided Neural Architecture Search. Retrieved from https://arni-institute.org/research-connectome-guided-neural-architecture-search. (2025).

4. AI for health

4.1 Background

The rapid development of AI technology has brought new opportunities to the healthcare field. From early disease screening and precision treatment planning to drug discovery and health management, AI applications are gradually covering every aspect of healthcare, but still face many technical and ethical challenges. Major issues include how to balance technological innovation with privacy protection, how to ensure the reliability and transparency of AI, and how to overcome barriers in interdisciplinary collaboration.

4.2 Recent advances

4.2.1 Disease diagnosis and prediction

AI has demonstrated great potential in early disease screening, diagnosis and prognosis analysis. Deep learning has surpassed radiologists in imaging diagnostics for diseases such as lung and breast cancer[1]. Med-PaLM2 passed the United States Medical Licensing Examination (USMLE) and can analyse conditions based on multimodal information such as X-rays[2]. The self-supervised pathology foundation model UNI, trained on more than 100 million histopathological images, can generalize across 108 cancer subtypes, significantly reducing reliance on data annotation[3].

4.2.2 Drug discovery

AI can rapidly screen candidate molecules of drug and optimize clinical trial processes, thereby reducing the drug development cycle and costs. The TNIK inhibitor INS018_055, the first AI-designed candidate drug using generative models for target discovery and molecular design[4], has entered Phase II clinical trials after promising results in Phase IIa trials.

4.2.3 Personalized medicine and health management

AI can integrate multi-omics data (such as genomics), lifestyle data, and individual treatment data to support personalized health management[5]. Smart wearable devices combined with AI-powered health assistants play a crucial role by analysing biological parameters, such as heart rate and blood glucose levels.

4.2.4 Public health and epidemic monitoring

AI has enhanced epidemic transmission modeling, vaccine distribution optimization, and the precise implementation of health policies[6,7]. During the COVID-19 pandemic, AI helped build real-time outbreak monitoring, transmission prediction, and identification of high-risk population, contributing to the evidence-based policies[8].

4.3 Key challenges and paths

4.3.1 The integration of multimodal data and professional expertise

Current AI systems are mostly limited to single-source data or scenarios. Integrating multimodal data with professional expertise from industry experts and build an interdisciplinary, multi-level ecosystem remains an important issue.

The paths to breakthroughs include:

- Develop a multimodal joint representation framework, leveraging structured knowledge from knowledge graphs to enhance the interpretability of fused representations.
- Build interactive systems that allow medical experts to annotate, correct and validate AI outputs.

4.3.2 Ensuring model fairness and credibility

Due to the heterogeneity of health data and sample biases, AI may introduce algorithmic discrimination in health predictions and diagnoses. Additionally, the lack of transparency in AI decision-making during health-related tasks undermines the trust of medical professionals and the public. Building AI technologies that ensure fairness and credibility is vital.

The paths to breakthroughs include:

- Develop explainable neural

network architectures to help medical professionals understand and trust the AI decision-making process.

- Strengthen algorithmic fairness constraints by establishing diverse data standards and cross-population evaluation metrics to ensure fair AI healthcare support across gender, ethnicity and region.

4.3.3 Data silos and privacy protection issues

Healthcare data is scattered across institutions and systems, forming data silos that are difficult to share and integrate effectively. Health data is highly sensitive, so to share and integrate, while ensuring privacy is an urgent problem.

The paths to breakthroughs include:

- Advance federated learning technologies to develop distributed model training without exchanging data.
- Enhance privacy protection technologies to provide technical support for cross-border health data collaboration.
- Build standardized testing benchmarks and certification systems for medical AI to promote interdisciplinary data and model sharing.

4.3.4 Ethical and regulatory challenges

The application of AI in healthcare involves ethical issues, making it difficult to define responsible parties. Current legal frameworks struggle to keep pace with fast-evolving technologies.

The paths to breakthroughs include:

- Establish ethical and legal frameworks specifically for medical AI, clarifying responsible parties, regulating data usage and model applications, and safeguarding patient rights.
- Introduce a 'sandbox regulation' mechanism, allowing innovative technologies to operate within defined limits while collecting clinical feedback to inform policy updates.

AI is transforming global healthcare. While its full potential is yet to be realized, the development of AI requires the coordinated progress in technology, ethics, regulation and application. Continued research and responsible innovation will not only drive technological progress but also profoundly benefit human health.

©Tanyajoy / iStock / Getty

References

1. McKinney, S.M. et al. International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89-94 (2020).
2. Karan Singhal, et al. Towards expert-level medical question answering with large language models. *arXiv preprint* arXiv: 2305.09617 (2023).
3. Chen, R.J. et al. Towards a general-purpose foundation model for computational pathology. *Nat. Med*. **30**, 850-862 (2024).
4. Ren, F. et al. A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nat. Biotechnol.* **43**, 63-75 (2025).
5. Liu, J. et al. Digital phenotyping from wearables using AI characterizes psychiatric disorders and identifies genetic associations. *Cell*/**188**,515-529 (2025).
6. Budd, J. et al. Digital technologies in the public-health response to COVID-19. *Nat Med.* **26**,1183-1192 (2020).
7. Syrowatka, A. et al. Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. *NPJ Digit Med.* **4**, 96 (2021).
8. Leung, K. et al. Quantifying the uncertainty of CovidSim. *Nat Comput Sci* **1**, 98–99 (2021).

5. AI-driven evolutionary studies

5.1 Background

Evolution is the foundational concept for understanding biodiversity. For decades, scientists have relied on traditional methods such as fossil analysis, gene sequencing and field observations to study evolution. However, the emergence of AI has fundamentally transformed these paradigms. AI's capacity to process massive datasets, identify complex patterns, and generate predictive models is unlocking new dimensions in the study of evolutionary mechanisms .

AI itself is evolving at an exponential pace. Its influence now extends across healthcare, transportation, communication, and beyond. A bidirectional interaction now exists between AI and evolution: AI accelerates biological evolution research, while AI-driven technological evolution inversely influences the evolutionary trajectories of both human and non-human species.

5.2 Recent advances

5.2.1 AI applications in biological evolution

5.2.1.1 Genomic analysis

Mutation detection: CNN-based algorithms precisely identify functional mutations in DNA sequences and predict their responses to selection pressures (e.g., adaptive or neutral mutations).

Phylogenetic reconstruction: AI-enhanced methods combining Bayesian inference and neighbor-joining methods improves evolutionary tree accuracy, resolving complex genetic relationships, such as speciation events.

5.2.1.2 Phenotypic analysis

Morphological traits: computer vision quantifies leaf venation patterns and 3D fossil structures, revealing morphology-environment coevolution dynamics.

Behavioral traits: wearable devices combined with machine learning algorithms analyse animal social behaviors (e.g., primate grooming), linking behavioral patterns to survival advantages.

5.2.1.3 Evolutionary simulation

Genetic algorithms: simulate the evolution antibiotic resistance in bacteria, optimizing parameter selection under selective pressures.

Agent-Based Models (ABM): construct predator-prey dynamic systems to analyse trait evolution and environmental adaptation mechanisms.

Model optimization: EVO1 implemented a 7B-parameter genomic foundation model processing 131kb DNA context at single-nucleotide resolution. Trained on 2.7M prokaryotic/phage genomes, it integrates central dogma modalities (DNA/RNA/protein) with evolutionary multiscale learning[1]. EVO 2 is a 7B/40B-parameter DNA foundation model trained on 9.3 trillion bases across life domains. With a 1M-token context window at single-nucleotide resolution, it generates biologically realistic sequences (mitochondrial, prokaryotic and eukaryotic), surpassing prior methods in naturalness[2].

5.2.2 AI-driven impacts on technological evolution

5.2.2.1 Healthcare and human evolution

Shifting Selection Pressures: AI-driven diagnostics enhance early disease intervention, reducing natural selection intensity on genetic defects (e.g., cancer susceptibility).

Human Augmentation: exoskeletons and cognitive-enhancing devices transcend biological limits, potentially reshaping competitive traits (e.g., learning efficiency).

5.2.2.2 Environmental interventions

Precision agriculture: AI-driven pesticide applications accelerate pest resistance evolution and alter soil microbial communities.

Robot Interactions: Agricultural robots and surveillance drones are influencing wildlife behavioral patterns (e.g., migration routes) and reshaping crop ecosystems.

Case studies include:

AlphaFold: Protein folding predictions reveal evolutionary conservation, accelerating drug target discovery.

Evolutionary Robots: MIT's genetic algorithm-based robots achieve environmental adaptability, mimicking natural selection mechanisms.

5.3. Key challenges and paths

5.3.1 Research biases

Data bias: overrepresentation of certain populations in genomic datasets can skews evolutionary inferences.

Algorithmic bias: incorrect assumptions in phylogenetic models may misguide biodiversity conservation strategies.

5.3.2 Evolutionary intervention risks

Gene editing: AI-assisted CRISPR technologies raise concerns over off-target mutations or ecological disruptions (e.g., gene drive organism proliferation).

Technological Inequality: Unequal access to cognitive-enhancing devices may exacerbate social disparities, influencing human trait evolution.

5.3.3 Governance frameworks

Ethical oversight: establish a Global Ethics Review Board for AI-driven evolutionary interventions.

Impact assessment: implement Technology Impact Lifecycle Assessment (TIA-LCA) protocols to evaluate long-term societal and ecological effects.

5.4. Conclusion and future directions

The integration of AI and evolutionary science marks a transformative research frontier. AI not only deepens our understanding of biological evolution but also reshapes cross-species evolutionary trajectories and technological pathways.

5.4.1 Scientific frontiers

Uncover adaptive evolution mechanisms (e.g., high-altitude adaptation) and gene-environment interactions.

Refine the construction of the 'Tree of Life' using AI-enhanced phylogenetic tools.

5.4.2 Technological innovations

Transform precision medicine, sustainable agriculture and ecological restoration. Furthermore, develop hybrid quantum-classical simulators for large-scale evolutionary modeling.

5.4.3 Societal imperatives

One of the goals is to address systemic biases in AI-driven evolutionary analyses. Therefore, it is important to ensure equitable access to technologies to prevent socioeconomic divides.

By harmonizing ethical governance with AI's analytical power, we can unlock new insights into Earth's biodiversity and foster technologies that support the coevolution of ecosystems and human civilization.

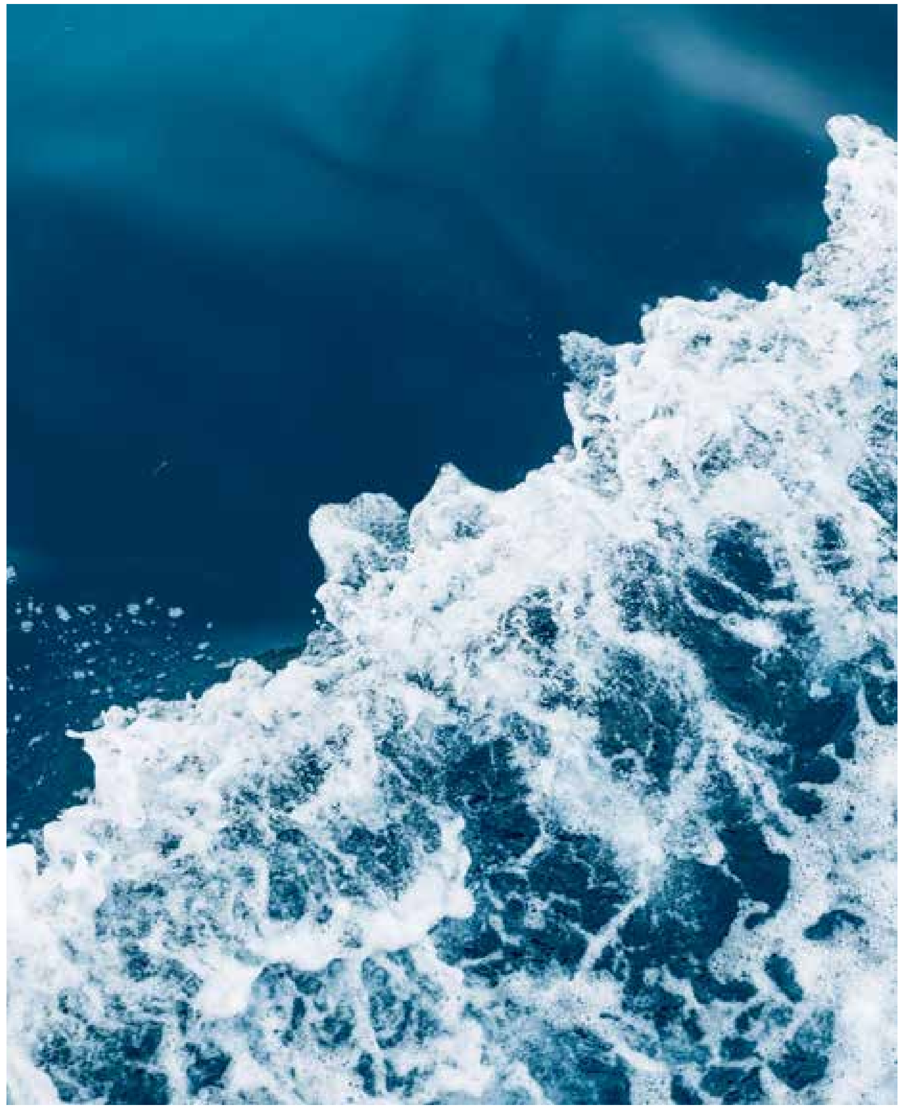
References

1. Nguyen, E. et al. Sequence modeling and design from molecular to genome scale with Evo. *Science* **386**, eado9336 (2024).
2. Garyk, B. et al. Genome modeling and design across all domains of life with Evo 2. doi: https://doi.org/10.1101/2025.02.18.638918 (2025).

Chapter 6

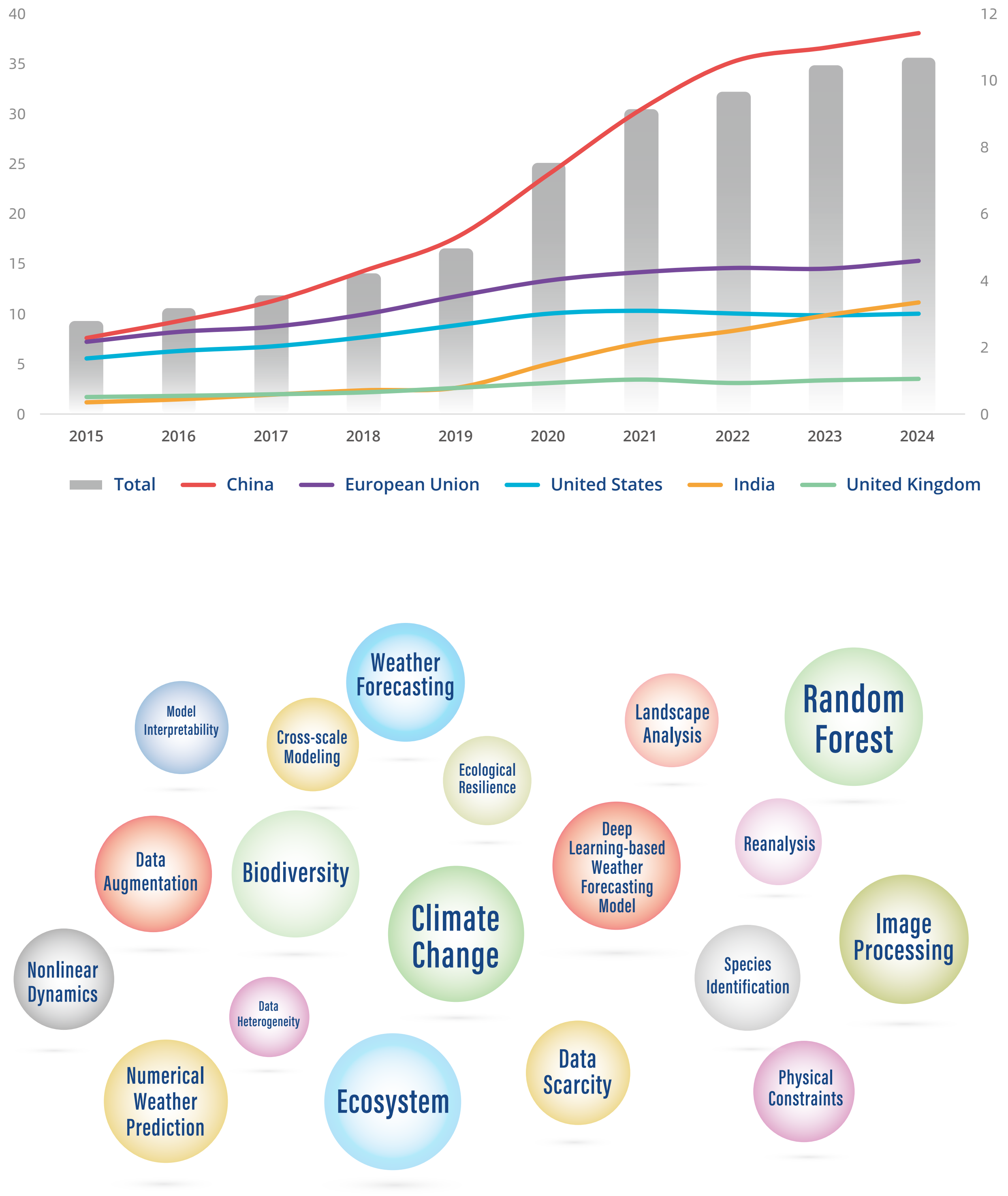
AI FOR EARTH AND ENVIRONMENTAL SCIENCES

Between 2015 and 2024, AI publications in the Earth and environmental sciences quadrupled, rising from 9,200 to 35,600 papers (Figure 6). China accounted for nearly half of global publications, far surpassing the output of the EU and the US. Meanwhile, India showed a strong upward trajectory, overtaking the US by 2024. Data analysis and the keyword cloud indicate highlight climate change, ecosystems and biodiversity as the most prominent research areas. Random forest methods and image processing techniques are widely used AI techniques, while multimodal and end-to-end AI models integrated with numerical forecasting have significantly improved the accuracy and efficiency of weather prediction. By incorporating physical constraints, employing data augmentation techniques, and applying interpretable models, AI is reshaping the landscape of Earth and environmental sciences.



©Mariia Siurtukova / Moment / Getty

FIGURE 6 | Earth and Environmental Sciences – Total AI Publications, National Trends (in thousands), and Keyword Cloud (2015–2024)



1. AI for weather forecasting

1.1 Background

Weather forecasting plays a crucial role in disaster mitigation, agricultural planning and enhancing societal resilience. Historically, weather predictions relied on empirical knowledge and statistical methods. This changed in the mid-20th century with the emergence of numerical weather prediction (NWP), which marked a transformative era in the field. Since the first successful numerical weather forecast in 1950 — solving the barotropic vorticity equation on ENIAC computer — marked a paradigm shift. Since then, the quiet revolution of NWP models has steadily improved forecast accuracy, earning NWP recognition as one of the most significant scientific achievements of the 20th century[1]. Today, leading operational centres have developed sophisticated numerical forecasting systems to generate daily forecasts.

NWP models consist of two primary components, one is the dynamical core which solves the governing equations of atmospheric evolution[2]. The other is physical parameterization schemes that approximate subgrid-scale processes such as radiation, convection, cloud microphysics, planetary boundary layer, and land-surface processes and interactions. However, these parameterizations inherit uncertainties arising from observational gaps, computational limitations and incomplete physical understanding — ultimately affecting model accuracy.

Despite significant advances in computational power, progress in NWP accuracy has plateaued. This stagnation stems from persistent reliance on empirical parameterizations and substantial computational costs and parallelizing models on modern supercomputers[3]. Meanwhile, advances in deep learning have emerged as a transformative paradigm for weather forecasting.

1.2 Recent advances

Since 2022, state-of-the-art models[4-8] have achieved medium-range forecast skill that comparable or superior to those of the high-resolution deterministic forecasts (HRES) by European Center for Medium-Range Weather Forecasts (ECMWF. Remarkably, these data-driven models can generate forecasts within seconds on a single GPU, bypassing the hours-long computations required by conventional NWP models on thousands of CPU cores on a supercomputer. Furthermore, operational centres like ECMWF and China Meteorological Administration (CMA) are already deploying deep learning-based weather forecasting. For instance, on June 18, 2024, CMA released three AI meteorological forecast models — Fengqing, Fenglei and Fengshun — designed for global medium-range forecasting, nowcasting and subseasonal-to-seasonal forecasting, respectively. Similarly, ECMWF began operational deployment of its Artificial Intelligence Forecasting System (AIFS) on February 25, 2025, running alongside its traditional physics-based models.

1.3 Key challenges and paths

Despite remarkable accomplishments, many challenges remain for deep learning-based weather forecasting models, which also present promising future research opportunities.

1.3.1 Poor extreme weather forecasts

Deep learning-based weather forecasts become increasingly blurry as the forecast lead time increases[9], because these models are typically trained to minimize loss functions such as mean squared error (MSE) or mean absolute error (MAE), which prioritize overall accuracy over preserving fine-scale details. While autoregressive, multi-step loss functions help mitigate the rapid accumulation of errors over long lead times, they often results in overly smooth forecasts. This issue is particularly pronounced for extreme weather events, such as heavy rainfall and strong winds — which have profound impacts, underscoring the urgent need to improve forecasts.

To address the challenge of blurry forecasts, probabilistic generative models have emerged as promising tools[10]. However, significant gaps remain in accurately forecasting extreme weather events. For instance, the intensity of tropical cyclones (TCs) is often substantially underestimated[5-7]. This limitation is partly due to the rarity and limited representation of extreme events. Additionally, the relatively smaller magnitudes of TC intensity in widely used training datasets like ERA5 constrain the ability of deep learning models. Retrieval-augmented generation (RAG) offers a potential solution by leveraging historical weather patterns that resemble the current state, enabling the recreation of past conditions as references to refine predictions. This approach is promising for improving the accuracy and robustness of extreme weather forecasts by addressing the challenges related to data scarcity. The following subsection discusses strategies for creating enhanced datasets to replace ERA5, with a focus on improving forecasts for extreme events like TCs.

1.3.2 Over-reliance on reanalysis data

Most deep learning weather forecasting models depend on ERA5 reanalysis data for training, as it is widely regarded as the most comprehensive and accurate global reanalysis archive[11]. However, ERA5 has several known biases. For instance, its precipitation often differs significantly from rain gauge observations or radar-based estimates[12]. Additionally, TC intensity data derived from ERA5 tends to be weaker than that from the International Best Track Archive for Climate Stewardship (IBTrACS) dataset, with higher mean sea level pressure and lower wind speed.

NWP models are well-suited for generating synthetic weather data to augment training datasets, thereby enhancing the robustness of deep learning-based forecasts[3]. For example, Ran et al.[13]proposed HR-Extreme, a comprehensive dataset that includes 3-km resolution extreme weather cases derived from the High-Resolution Rapid Refresh (HRRR). Furthermore, NWP models can be used to downscale global weather forecasts, producing more accurate TC datasets. They also enable the generation of unseen weather conditions, which are critical for improving the generalization capabilities of deep learning models.

1.3.3 Lack of physical constraints

Most deep learning weather forecasting models are purely data-driven. While deep learning models can emulate weather patterns when trained on reanalysis data, there is no guarantee that their predictions are physically consistent (e.g., conserving mass or energy). For example, deep learning models can produce unphysical outputs, such as negative humidity[14], and may lack fidelity in certain process-based verification aspects, such as butterfly effects[15] and geostrophic balance[9]. These limitations often stem from the models' failure to incorporate physical relationships.

The integration of physical constraints into deep learning-based weather forecasting models is still in its early stages. The hybrid approach, combining deep learning models with physical laws, holds great potential. Partial differential equations (PDEs) can be utilized to describe dynamic processes, while ambiguous processes can be modeled using AI approaches. Recently, Kochkov et al. introduced the NeuralGCM model[16], which integrates a differentiable dynamical core for solving discretized governing equations without having explicit physical constraints. Incorporating physical constraints also holds significant promise for improving the out-of-distribution generalization capabilities of deep learning models.

1.3.4 Limited spatial and temporal resolution

Most state-of-the-art deep learning weather forecasting models generate forecasts at a spatial resolution of 0.25 ° (corresponding to 721 × 1440 grid points). Two primary challenges prevent deep learning-based global weather forecasting models from achieving higher spatial resolution. First, the lack of higher-resolution reanalysis or analysis data constrains the ability of deep learning models to generate higher-resolution forecasts. Currently, the highest-resolution global analysis data is ECMWF HRES, which has a spatial resolution to 0.09 ° since March 8, 2016. However, this dataset is much smaller compared to the widely used ERA5 reanalysis data[10], which is commonly used for developing deep learning weather models. Second, higher spatial resolution increases computational complexity, requiring significant investment to train higher-resolution models.

NWP models such as Weather Research and Forecasting Model (WRF) can provide high-resolution training data for training deep learning based weather forecasting models with higher spatial resolutions.

In addition, most models produce forecasts at 6-hour intervals, which lack the temporal resolution required for many applications. The preference for 6-hour over 1-hour forecasts arises from error accumulation in autoregressive models[7], a long-standing challenge in deep learning-based weather forecasting. Pangu-Weather[6] employs a hierarchical temporal aggregation strategy for 1-hour forecasts but suffers from continuity issues due to varying iteration requirements across consecutive steps.

1.3.5 Atmosphere-only modeling

Most deep learning weather forecasting models only use atmospheric data to predict future weather conditions. However, the adoption of an Earth system modeling by integrating the atmosphere, oceans, land, and cryosphere could further advance weather forecasting by capturing the interconnected and interdependent processes that drive weather and climate. For example, ocean-atmosphere interactions, such as El Niño and La Niña, significantly impact global weather systems. These phenomena are driven by changes in sea surface temperatures and heat exchange, which are not captured by atmospheric models alone. Extreme weather events, such as TCs, floods and heatwaves, often result from interactions between multiple Earth system components.

Coupling atmospheric models with ocean and land models allows for better representation of sea surface temperatures and land surface conditions, which are critical for predicting TCs and floods. In its strategy for 2025 to 2034, ECMWF identifies improving monitoring and predictions of the Earth system as one of its top priorities.

References

1. Bauer, P. et al. The quiet revolution of numerical weather prediction. *Nature* **525**, 47-55 (2015).
2. Kalnay, E. Atmospheric modeling, data assimilation and predictability. *Cambridge university press* (2003).
3. Bauer, P. What if? numerical weather prediction at the crossroads. *JEMS* **1**, 1-12 (2024).
4. Pathak, J. et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. (2022).
5. Bi, K. et al. Accurate medium-range global weather forecasting with 3d neural networks. *Nature* **619**, 533–538 (2023).
6. Lam, R. et al. Learning skillful medium-range global weather forecasting. *Science* **382**, 1416-1421 (2023).
7. Chen, L. et al. Fuxi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj Clim. Atmos. Sci.* 1-11 (2023).
8. Lang, S. et al. AIFS — ECMWF's data-driven forecasting system. *arXiv preprint* **arXiv:2406.01465** (2024).

9. Bonavita, M. On some limitations of current machine learning weather prediction models. *Geophys. Res. Lett.* **51**, 2023-107377 (2024).
10. Price, I. et al. Probabilistic weather forecasting with machine learning. *Nature* **637**, 84-90 (2025).
11. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.***146**, 1999-2049 (2020).
12. Lavers, D.A. et al. An evaluation of era5 precipitation for climate monitoring. *Q. J. R. Meteorol. Soc.* **148**, 3152-3165 (2022).
13. Ran, N. et al. HR-Extreme: A high-resolution dataset for extreme weather fore- casting. *arXiv preprint* **arXiv:2409.18885** (2025).
14. Schreck, J. et al. Community Research Earth Digital Intelligence Twin (CREDIT) (2024).
15. Selz, T. et al. Can artificial intelligence-based weather predic-tion models simulate the butterfly effect? *Geophys. Res. Lett.* **50**, 2023-105747 (2023).
16. Kochkov, D. et al. Neural general circulation models for weather and climate. *Nature* **632**, 1060-1066 (2024).

2. AI for environmental science

2.1 Background

In the face of climate change, resource scarcity, environmental pollution and ecological degradation, environmental science plays a pivotal role in elucidating natural mechanisms, formulating governance strategies and ensuring sustainable development. It also provides theoretical foundations for addressing global issues like pollution and ecosystem degradation.

AI enables the integration of multi-source data—including remote sensing imagery, field observations and laboratory measurements—to enhance the spatiotemporal prediction accuracy of environmental factors. AI facilitates rapid detection and precise prediction of environmental changes, driving transformative advances in environmental science.

Moreover, AI holds great promise for the construction of data-driven, multi-scale and multi-process coupled environmental models capable of simulating complex systems and quantifying uncertainties. AI will continue to advance intelligent environmental monitoring, governance and management, offering strong support for ecological conservation and the achievement of sustainable development goals.

2.2 Recent advances

AI technology has achieved remarkable progress in environmental science, with a rapidly expanding range of application.

First, multi-source data fusion techniques, which integrate satellite remote sensing, ground-based observations, geographic information and statistical datasets, have significantly enhanced the precision of spatiotemporal variation analysis for environmental factors, as well as the timeliness and accuracy of environmental monitoring and early warning[1,2]. AI-driven automated processing systems enable rapid estimation of disaster outbreaks and propagation, such as earthquakes, wildfires, floods and volcanic eruptions, greatly improving emergency response efficiency[3,4]. Automated environmental monitoring networks, integrated with machine learning algorithms, enable real-time, high-efficiency surveillance of pollution, biodiversity changes and flood disasters. These systems effectively reduce labour costs while improving data quality[5].

Second, AI applications in environmental system modeling have rapidly expanded. The integration of deep learning and graph neural networks has significantly improved the simulation accuracy of complex Earth system models (e.g., atmospheric, oceanic, and terrestrial systems) and multi-scale climate processes[6]. These improvements enhance the understanding of nonlinear relationships and underlying mechanisms in ecosystems, climate systems, atmospheric systems and socio-economic systems, providing robust scientific support for environmental policy formulation and impact assessment[7-10]. Meanwhile, interdisciplinary collaborations have further driven the use of AI in disaster prediction, species conservation, and sustainable urban planning.

Additionally, LLMs such as ChatGPT and DeepSeek become increasingly important in environmental science[11]. LLMs enable rapid processing of massive environmental datasets, assisting researchers in extracting valuable insights while enhancing the accuracy of data analysis. They also support efficient scientific writing by generating high-quality textual content and offering intelligent decision-making tools for policy development and evaluation. Through natural language processing, LLMs facilitate cross-domain knowledge sharing and dissemination, helping to accelerate progress in environmental research,.

2.3 Key challenges and paths

The integration of AI and environmental science is advancing the field into a new stage, with core directions focusing on three dimensions: data integration, mechanism fusion and intelligent decision-making.

In the domain of complex environmental system modeling, integrating multi-source data from satellite remote sensing, ground-based sensors and socio-economic activities remains a significant challenge. Substantial discrepancies in spatiotemporal resolution and sampling standards hinder the alignment of heterogeneous datasets, the construction of environmental knowledge graphs, and the embedding of physical laws into neural network constraints.

Predicting environmental tipping points faces challenges due to abrupt transitions characteristic of nonlinear systems. Traditional models struggle to capture precursor signals of critical shifts driven by multi-factor interactions, while AI techniques optimized for minimizing overall errors, tend to neglect predictions of extreme events and tipping points.

The conflict between ecological conservation and economic development underscores the need for innovative intelligent decision-making technologies. The effectiveness of multi-agent reinforcement learning frameworks are critical to transforming policy-making models[12].

The key to AI technology breakthroughs relies on the standardization and open access of environmental data, the deep fusion of environmental mechanisms with AI architectures, and the holistic integration of natural and socio-economic processes. Researchers should prioritize establishing unified standards for data acquisition, processing and storage, as well as cross-institutional and interdisciplinary open-data platforms. These platforms should integrate the real-time data from satellites, ground sensors, and drones; statistical data from human activities; and theoretical data from laboratories. Intelligent preprocessing techniques should be used to remove noise and outliers. By incorporating attention mechanisms, locally interpretable models and sensitivity analysis tools, researchers can reveal the internal information processing of AI models.

Meanwhile, physics-guided deep learning models, developed through principles of environmental science, ecology and socioeconomics, should ensure both accuracy and scientific interpretability. Breakthroughs in intelligent decision-making require the integration of multi-source data, traditional numerical models, data-driven models and LLMs. Such integration should support process for problem identification, research execution, decision-making and outcome evaluation.

References

1. Li, Z. et al. Learning spatiotemporal dynamics with a pretrained generative model. *Nat. Mach. Intell.* **6**, 1566-1579 (2024).
2. Gibson, P. B. et al. Training machine learning models on climate model output yields skillful interpretable seasonal precipitation forecasts. *Commun. Earth Environ.* **2**, 159 (2021).
3. Buster, G. et al. High-resolution meteorology with climate change impacts from global climate model data using generative machine learning. *Nat. Energy* **9**, 894-906 (2024).
4. Camps-Valls, G. et al. Artificial intelligence for modeling and understanding extreme weather and climate events. *Nat. Commun.* **16**, 1919 (2025).
5. Zhu, Q. et al. Building a machine learning surrogate model for wildfire activities within a global Earth system model. *Geosci. Model Dev.* **15**, 1899-1911 (2022).
6. Bodnar, C. et al. A foundation model for the earth system. *arXiv e-prints*, **arXiv:2405.13063** (2024).
7. Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195-204 (2019).
8. Gettelman, A. et al. The future of Earth system prediction: Advances in model-data fusion. *Sci. Adv.* **8**, eabn3488 (2022).
9. Eyring, V. et al. Pushing the frontiers in climate modelling and analysis with machine learning. *Nat. Clim. Change* **14**, 916-928 (2024).
10. Mansfield, L. A. et al. Predicting global patterns of long-term climate change from short-term simulations using machine learning. *npj Clim. Atmos. Sci.* **3**, 44 (2020).
11. Zhu, J.-J. et al. ChatGPT and Environmental Research. *Environ. Sci. Technol.* **57**, 17667-17670 (2023).
12. Reichstein, M. et al. Early warning of complex climate risk with integrated artificial intelligence. *Nat. Commun.* **16**, 2564 (2025).

3. AI for ecological science

3.1 Background

Ecological science aims to understand the structure, function and dynamic patterns of ecosystems, playing a critical role in addressing global challenges such as biodiversity loss and climate change. These crises often arise from nonlinear dynamic imbalances triggered by disturbances in complex systems, making them highly unpredictable. Recent breakthroughs in AI have revolutionized ecological research by providing tools to quantify and observe ecological phenomena that are difficult to capture through traditional methods, thereby enabling the development of more accurate predictive models.

Traditional ecological modeling has long been constrained by the complexity of ecological systems. However, the deep integration of AI is breaking through these limitations. Techniques such as machine learning and complex system simulation, are offering new paradigms for modeling ecological processes and interactions. This integration is fostering bidirectional innovation: ecological challenges are shaping the development of next-generation AI algorithms, while AI's computational power accelerates the discovery of ecological patterns.

This co-evolutionary mechanism lays a methodological foundation for the deep integration of ecological science and AI.

3.2 Recent advances

3.2.1 Studies on species distribution and ecological relationship

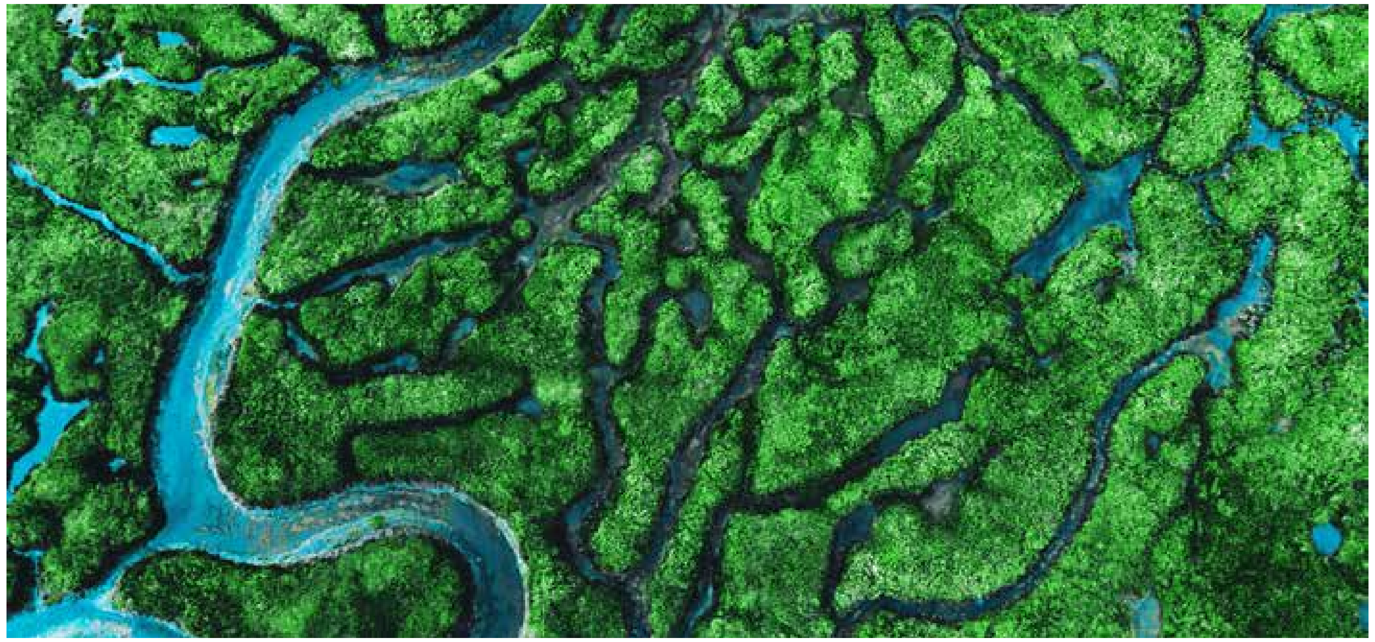
Through AI algorithms, researchers can analyse vast amounts of ecological data to reveal the complex relationships between landscape patterns and species distribution. This improves understanding of the habitat requirements of species and provides a scientific basis for biodiversity conservation.

3.2.2 Multi-source data fusion monitoring

Using AI technology, researchers can integrate data from diverse sources — such as optical, radar sensing — to generate high-resolution land cover change maps. This allows for the accurate identification of urban expansion areas, fallow agricultural and grassland distribution, as well as trends in forest loss and gain, providing real-time and accurate data support for land resource management.

3.2.3 Simulation experiments for ecological process

With models such as Long Short-Term Memory (LSTM) networks in deep learning, simulation experiments of ecological processes can be conducted. For example, by combining reinforcement learning with Multilayer Perceptron (MLP) neural networks and Markov chain analysis, researchers can simulate the long-term effects of floods on agricultural ecosystem services, and predict changes in urban expansion and green infrastructure deployment. These simulations provide valuable decision-making support for post-disaster land management.



©Artur Debat / Moment / Getty

3.2.4 Knowledge-guided machine learning (KGML)

KGML integrates scientific knowledge into the infrastructure of machine learning algorithms, enabling models to make predictions that better conform to physical laws[1].

3.2.5 Integration of deep learning and mechanistic models

The integration of deep learning with traditional ecological mechanistic models offers a promising pathway for improving both predictive accuracy and model interpretability. Deep learning models can handle complex nonlinear relationships, while mechanistic models have explicit physical significance and interpretability. The fusion of these approaches allows for a more comprehensive understanding of ecosystems.

3.3 Key challenges and paths

3.3.1 Model interpretability

Many AI models, such as deep learning models, are often regarded as 'black-box' systems, due to the difficulty of explaining their internal mechanisms[2]. This lack of interpretability hinders researchers from validating model outputs and limits their utility in supporting reliable ecological management decisions.

The paths to breakthroughs lie in incorporating domain knowledge. This means incorporating fundamental ecological principles — such as energy flow and nutrient cycling — into AI model design, which allows the model to reason and predict based on these principles.

3.3.2 Data quality and scarcity

Ecological data often suffer from issues such as inconsistent quality, missing values and noise interference. Additionally, the difficulty of acquiring certain ecological data leads to scarcity, which limits the availability of training samples and affects model generalizability and reliability.

The paths to breakthroughs include multi-source data integration. By leveraging complementary multi-source data, issues of missing data and noise can be mitigated, providing richer and more reliable inputs for AI models.

3.3.3 Cross-scale and cross-regional generalization

Current AI models often lack of generalization across scales and regions, as models trained in one context cannot be directly applied to others. This restricts their applicability and effectiveness 3.

The paths to breakthroughs include data augmentation and transfer learning. Models pre-trained in fields like image recognition can be adapted via transfer learning for ecological remote sensing image classification and analysis, enhancing model performance even with limited data[3].

3.4 Conclusion and outlook

Future research should focus on addressing key challenges such as data scarcity, model interpretability and cross-scale modeling to deepen the integration of AI and ecological science. The complexity and resilience of ecosystems offer new inspiration for AI development. Strengthening interdisciplinary research between these fields will drive mutual progress and innovation[4].

References

1. Read, J. S. et al. Process-guided deep learning predictions of lake water temperature. *Water Resour. Res*. **55**, 9173-9190 (2019).
2. George, LWP. et al. An outlook for deep learning in ecosystem science. *Ecosystems* **25**, 1700-718 (2022).
3. Yu, Z. et al. Machine learning for ecological analysis. *Chem. Eng. J* **507**, 160780 (2025).
4. Barbara, AH. et al. A synergistic future for AI and ecology. *PNAS* **120**, e2220283120 (2023).

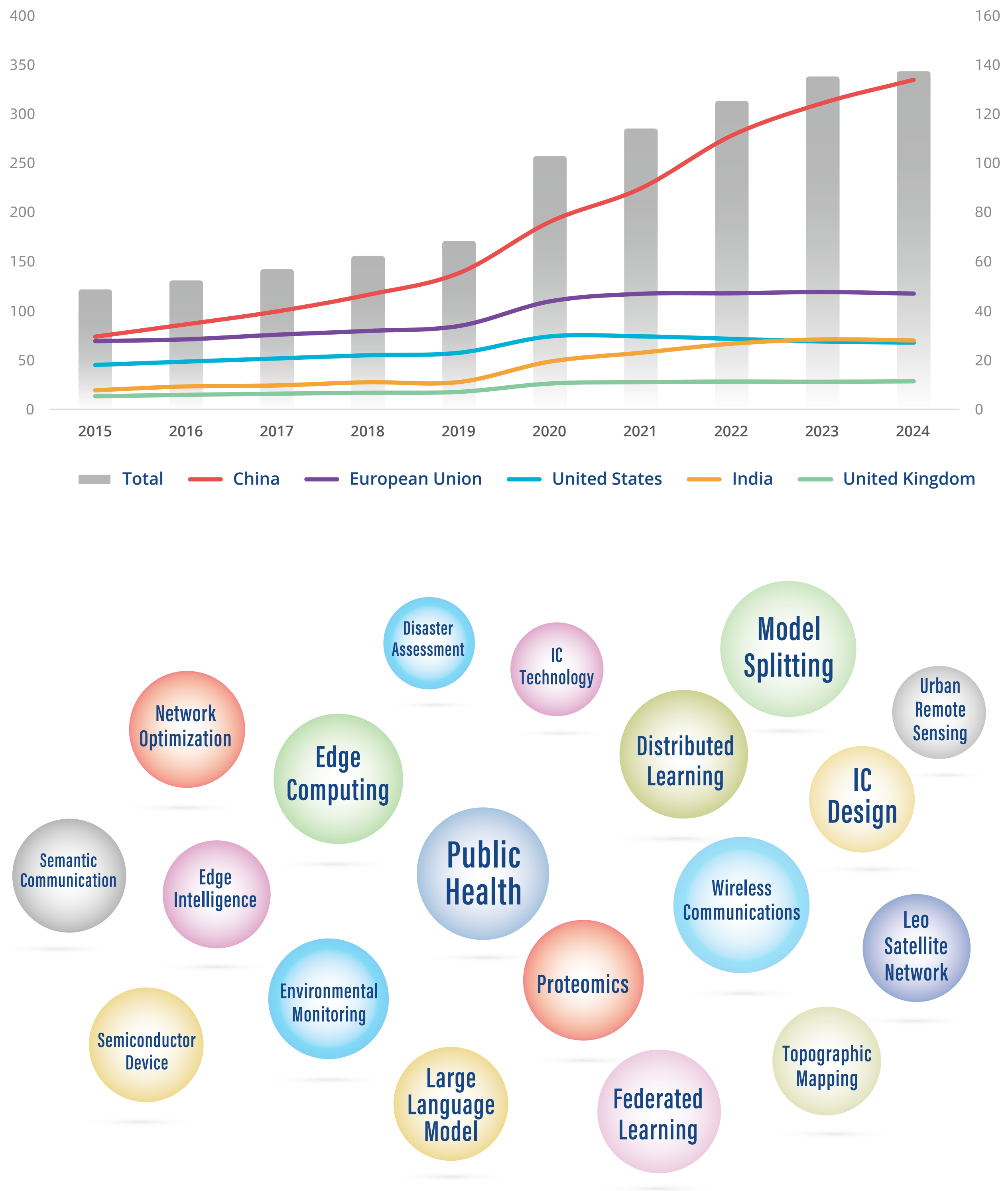
Chapter 7

AI FOR ENGINEERING

©Xuanyu Han / Moment / Getty

Between 2015 and 2024, global AI publications in engineering sciences nearly tripled, rising from 121,600 to 343,000 papers. China maintains a decisive lead, surpassing both the US and the EU, while India has rapidly caught up — overtaking the US by 2024 (Figure 7). Data analysis and the keyword cloud highlight wireless communications, network optimization, semiconductor design and remote sensing as the key areas of focus. AI methodologies, such as large language models and federated learning, have become indispensable in engineering. Meanwhile, cutting-edge technologies — such as semantic communication, edge intelligence and model splitting — are embedding AI into the backbone of next-generation infrastructure, driving a deep integration of ecological sustainability and system resilience.

FIGURE 7 | Engineering – Total AI Publications, National Trends (in thousands), and Keyword Cloud (2015–2024)



1. AI for communications

1.1 Background

With the rapid development of information technology, the integration of AI into communication technologies has become a key driver of next-generation innovation. AI is recognized as the core technology in the architectures of sixth-generation mobile networks (6G) and sixth-generation fixed networks (F6G)[1,2]. The AI-RAN Alliance, established by several leading industry companies, emphasizes intelligent radio access networks (RAN) as a primary direction for development[3].

This integration holds far-reaching implications. From the technical perspective, AI empowers communication systems with self-optimizing capabilities, enhancing reliability, reducing latency, and improving energy efficiency. Economically, it enables new business models and application scenarios, driving economic growth. Socially, it supports services such as smart cities, intelligent transportation, and telemedicine, substantially enhancing quality of life.

The fusion of AI with communications technology is not merely an inevitable trend in communications development but also a crucial path toward realizing the vision of future communications.

1.2 Recent advances

Rapid advances in AI are profoundly reshaping the theoretical frameworks, experimental methods, and modeling approaches of modern communication systems. AI is driving a significant shift from traditional Shannon-based communication paradigms towards new models such as semantic communication. Mathematical frameworks such as semantic entropy, semantic mutual information and semantic channel capacity have expanded the boundaries of classical information theory, providing novel theoretical foundations for designing highly efficient and low-latency communication systems.

Techniques involving deep learning, multimodal data processing and intelligent decision-making have led to dramatic improvements — exceeding tenfold gains in network capacity, coverage and energy efficiency.

Simultaneously, international standards organizations such as 3GPP actively promote the deployment of AI in wireless networks. Utilizing AI models for beam management, channel prediction, and network scheduling significantly improves system responsiveness and stability, forming a strong foundation for future 6G networks.

Technologies such as distributed AI model transmission, intelligent control, and adaptive decision-making enable optimal network resource allocation and significant energy consumption reductions.

Cross-disciplinary integration between communications and AI is accelerating the development of advanced AI hardware accelerators and algorithmic platforms. Next-generation AI accelerators and specialized chips, facilitated by ultra-wideband interconnection networks, enable real-time training and inference of complex neural networks, further enhancing communication network intelligence.

Overall, the integration of AI is propelling the communications field toward greater efficiency, enhanced intelligence and lower energy consumption.

1.3 Key challenges and paths

1.3.1 AI for communications (AI4Com)

While AI-driven communications offer immense potential to enhance system performance and efficiency, the strong dependence of deep neural networks on vast training data volumes poses a significant challenge.

To balance data dependency with model interpretability, technological breakthroughs are required across data, algorithms, systems and hardware domains. Incorporating physical models can enhance interpretability, reduce computational latency, and avoid local optima. Progress in these areas could enable end-to-end optimization, from the physical layer to the network layer, driving the stable development of 6G and subsequent communication systems.

1.3.2 Communications for AI computation (Com4AI)

Communication systems are emerging as crucial support for computational networks, especially in distributed AI training. A major challenge lies in understanding the network architectures and collaborative mechanisms underlying communications-enhanced AI computation.

Research at the network layer focuses on intelligent scheduling and flow management optimization using software-defined networking (SDN) to improve communication efficiency. At the physical layer, integrating optical computing and optical interconnection technologies provides a promising approach to overcome the limitations of traditional electronic systems, especially in terms of power consumption and speed.

1.3.3 Deep integration for native intelligent communication (COM-AI)

The deep integration of AI with communications is transitioning from tool-assisted to native systems, forming a closed-loop framework of perception, transmission, and decision-making. This transformation is catalyzing innovative paradigms such as semantic communication, brain-computer interfaces, and large-scale communication models, raising numerous challenges.

Advances in semantic modeling, computational architectures, optimization strategies, and resource scheduling are crucial to overcome these challenges. Exploring large-scale models specifically tailored for communication systems will drive the evolution of communication systems from traditional optimization toward native intelligence.

References

1. Tong, W. et al. 6G: The Next Horizon. *Cambridge Univ. Press* (2021).
2. Uzunidis, D. et al. A vision of 6th generation of fixed networks (F6G): challenges and proposed directions. *Telecom*. **4**, 758-815(2024).
3. AI-RAN Alliance. AI-RAN Alliance Vision and Mission White Paper[EB/OL]. (2024).

2. AI for remote sensing

2.1 Background

In recent years, rapid advances in AI — particularly in large-scale foundational models — have driven transformative innovations in many domains. In the field of remote sensing, foundational models have gradually emerged, aiming to build unified large-scale models for diverse remote sensing data and image interpretation tasks.

The processing framework includes multimodal remote sensing big data, remote sensing foundational models, and Earth observation applications. These models can process multi-source, multi-resolution and multi-band remote sensing images, serving tasks such as scene classification, object detection, tracking, image segmentation and change detection.

Integrating AI technologies with remote sensing science enhances image processing accuracy and multitasking capabilities, accommodates the growing complexity of remote sensing data analysis and driving the intelligent transformation of remote sensing systems.

2.2 Recent advances

2.2.1 Vision foundation models for multitemporal and multimodal remote sensing data

Vision foundation models that integrate multitemporal and multimodal remote sensing data enable dynamic monitoring of surface features by combining data from different time points and sensors. SatMAE[1], based on the Vision Transformer, performs pretraining on multispectral and synthetic aperture radar (SAR) images, improving capabilities in land cover change monitoring. RingMo[2] employs masked autoencoding on Vision Transformer to construct a global multitemporal dataset, demonstrating the potential of self-supervised pretraining.

2.2.2 Vision-language models for cross-modal remote sensing image interpretation

Vision-language models for cross-modal remote sensing image interpretation drive transformative changes in remote sensing intelligent interpretation through interactions between natural language and visual prompts. GeoChat[3] implements a versatile vision-language model capable of complex tasks such as image-level, region-level and localization analyses. EarthGPT[4] bridges the gap between cross-modal understanding and visual reasoning by constructing a million-scale multimodal dataset, supporting unified comprehension of optical, SAR, infrared, and other multi-sensor images for tasks like scene classification, image captioning and object detection.

2.2.3 Physics-inspired diffusion models for remote sensing image generation

Diffusion models[5] exhibit significant potential in remote sensing image generation, producing high-quality images through iterative denoising process. By embedding physical mechanisms into diffusion models, the generation process incorporates the physical characteristics of remote sensing images. DiffusionSat integrates spatiotemporal features to optimize spatial consistency. HSIGene[6] incorporate spectral properties to enhance the attention mechanism of diffusion models, effectively capturing spatial-spectral features.

2.3 Key challenges and paths

2.3.1 Scaling laws between remote sensing big data and foundational model parameters

The fusion of multimodal remote sensing data and the scaling laws of large models, along with the modeling of multi-source data distributions, have become a focal point of current research.

However, the scaling laws for remote sensing visual models remain unclear[7], systematic research is needed to optimize large model construction. Therefore, it is important to establish a theoretical framework and empirical system for foundational models based on standardized multitemporal and multimodal remote sensing big data. Through systematic experiments, optimize model topology and train models at varying parameter scales to explore evolutionary patterns between data scale and model parameters. This will reveal the intrinsic dimensionality of data manifolds and critical phase transition thresholds of model capacity, ultimately constructing a mathematical model.

2.3.2 Self-supervised pretraining algorithms and efficient post-training techniques for remote sensing data

Exploring efficient architectures for remote sensing data fusion and cross-modal learning strategies on self-supervised learning and supervised learning is essential to improve data alignment and representation capabilities.

Self-supervised pretraining and efficient post-training techniques for remote sensing data aim to reduce reliance on large annotated datasets while improving model generalization and training efficiency. In self-supervised pretraining, techniques such as image reconstruction, generation and context prediction are employed to build high-level representations[8]. Post-training focuses on efficient fine-tuning with limited annotations to enhance accuracy and adaptability for remote sensing tasks.

2.3.3 Reinforcement learning and reward feedback mechanisms for remote sensing visual tasks

Integrating reinforcement learning and reward feedback mechanisms with data-driven and physical-informed approaches is crucial to advancing intelligent remote sensing system. By introducing reinforcement learning and reward mechanisms[9], dynamic perception and adaptive decision strategies can transcend static modeling.

A key development direction involves building virtual-real training environments, combining physical radiation models with real satellite data to optimize models. In addition, designing multi-agent collaborative interpretation systems to enhance drone and satellite monitoring, developing task-adaptive reward mechanisms, and integrating geospatial-temporal priors to improve long-term task exploration efficiency are collectively contributing to a more intelligent remote sensing ecosystem.

References

1. Cong, Y. et al. SatMAE: Pre-training transformers for temporal and multi-spectral satellite imagery. *NeurIPS* **35**, 197-211 (2022).
2. Sun, X. et al. RingMo: A remote sensing foundation model with masked image modeling. *TGRS* **61**, 1-22 (2022).
3. Kuckreja, K. et al. GeoChat: Grounded large vision-language model for remote sensing. Proceedings of the *IEEE/CVF CVPR*. 27831-27840 (2024).
4. Zhang, W. et al. EarthGPT: A universal multi-modal large language model for multi-sensor image comprehension in remote sensing domain. *TGRS*. (2024).
5. Khanna, S. et al. DiffusionSat: A generative foundation model for satellite imagery. *ICLR*. (2024).
6. Pang, L. et al. HSIGen: A foundation model for hyperspectral image generation. *arXiv preprint* **arXiv: 2409.12470** (2024).
7. Kaplan, J. et al. Scaling laws for neural language models. *arXiv preprint* **arXiv: 2001. 08361** (2020).
8. Chen, T. et al. A simple framework for contrastive learning of visual representations. *ICML* 1597-1607 (2020).
9. Chen, J. et al. A reinforcement learning framework for scattering feature extraction and SAR image interpretation. *TGRS*. (2024).

3. AI for microelectronics

3.1 Background

As microelectronics devices approach their physical and material limits, the industry is facing critical challenges. First of all, traditional silicon-based materials are reaching performance bottlenecks due to quantum effects, thermal effects and current leakage, all of which hinder device improvement, particularly in terms of energy efficiency.

Secondly, the complexity of semiconductor manufacturing processes continues to rise, making it urgent to optimize process parameters to enhance yield and productivity. In addition, the scale and complexity of circuit design are increasing rapidly. Traditional design methods struggle to meet the demands for modern high-performance, low-power chips.

The emergence of AI and ML technologies brings new opportunities to the field of microelectronics. These technologies can efficiently handle large-scale data, optimize complex design and manufacturing processes, and uncover hidden patterns and optimization paths through data-driven approaches. They offer powerful tools to tackle the multi-dimensional and highly complex challenges that traditional methods cannot handle effectively.

3.2 Recent advances

The application of AI technologies offers new perspectives in materials science. Generative models can be used to create new semiconductor materials [1,2]. Machine learning models can predict key properties such as stability and band structure[3]. AI models can also directly generate materials that meet desired performance criteria[3]. AI-based high-throughput screening helps identify promising candidate materials, guiding experimental design and validation[4].

AI also shows great potential in semiconductor device modeling and structural discovery. Neural networks and Bayesian optimization can automate the calibration of device models, improving the accuracy of simulation results[5].

In semiconductor device modeling, generative models can produce new data samples from limited experimental data, expanding training datasets and enhancing model performance[6]. By combining evolutionary algorithms with machine learning, researchers can efficiently search for and optimize semiconductor material structures[7].

Semiconductor manufacturing involves complex process steps and equipment. Virtual metrology leverages machine learning models to predict potential quality issues during production[8]. Deep neural networks have been used for task allocation to equipment, improving manufacturing efficiency and reducing production cycles[9]. Machine learning has also been applied to equipment health prediction and fault detection, allowing early identification of issues and avoiding production interruptions[10]. Optimization techniques such as Bayesian optimization can minimize the number of experiments needed to find the best process parameters[11].

AI technologies are widely used in circuit design and electronic design automation (EDA). Neural networks are applied in performance modeling of RF/analog circuits and congestion analysis in digital IC design[12]. Techniques like Bayesian optimization and reinforcement learning are used for analog circuit optimization[13] and IC layout design. LLMs can automatically generate hardware description language (HDL) code, automate design processes based on user requirements, adjust analog circuit design parameters, and modify circuit topologies[14].

3.3 Key challenges and paths

Generative approaches, explainability, and the application of LLMs marks key frontiers in applying AI to microelectronics. Generative AI methods (such as GANs, VAEs, and diffusion models) can be used to generate novel materials and circuit structures, and their performance can be validated using methods like density functional theory (DFT) and molecular dynamics, thereby significantly improving design efficiency. However, challenges remain in terms of controllability, stability, and computational cost. The issue of model explainability limits AI applications in high-reliability scenarios. Techniques such as feature visualization, attention mechanisms, and decision trees can reveal the logic of models, helping engineers understand key influencing factors.

Most existing generative approaches are data-driven, whereas the microelectronics field relies heavily on physical laws. Combining physics-informed neural networks helps ensure that generated results comply with basic physical constraints. To enhance model explainability, visual analytics, symbolic AI and knowledge graphs can be integrated to help engineers understand AI decision logic. For LLM applications, incorporating knowledge graphs to generate fine-tuning data can improve both reasoning capabilities and domain-specific accuracy. Techniques like retrieval-augmented generation and formal verification can ensure the reliability of LLMs in complex design tasks. Meanwhile, building multi-agent systems can enable full intelligence across materials, devices, processes, and circuits.

References

1. Mannodi-Kanakkithodi, A. A guide to discovering next-generation semiconductor materials using atomistic simulations and machine learning. *Comput. Mater. Sci*. (2024).
2. Baird, S. G. et al. Data-driven materials discovery and synthesis using machine learning methods. *arXiv preprint* **arXiv:2202.02380v2** (2022).
3. Sorkun, M. C. et al. An artificial intelligence-aided virtual screening recipe for two-dimensional materials discovery. *npj Comput. Mater.* **6**, 106 (2020).
4. Pyzer-Knapp, E. O. et al. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *npj Comput. Mater.* **8**, 84 (2022).
5. Jeong, C. et al. Bridging TCAD and AI: Its Application to Semiconductor Design, *IEEE T. Electron. Dev*. (2021).
6. Wang, Z. et al. Improving Semiconductor Device Modeling for Electronic Design Automation by Machine Learning Techniques. *IEEE T. Electron. Dev*. (2024)
7. Choubisa, H. et al. Interpretable discovery of semiconductors with machine learning. *npj Comput. Mater.* (2023)
8. Tin, T. C. et al. Virtual Metrology in Semiconductor Fabrication Foundry Using Deep Learning Neural Networks. *IEEE Access* (2022).
9. Lee, Y.H. et al. Deep reinforcement learning based scheduling within production plan in semiconductor fabrication. *Expert Syst. Appl*. **191**, 116222 (2022).
10. Huang, A. C. et al. A survey on machine and deep learning in semiconductor industry: methods, opportunities, and challenges. *Cluster Computing*, (2023).
11. Kanarik, K. J. et al. Human–machine collaboration for improving semiconductor process development. *Nature* **616**, 707–711 (2023).
12. Min, K. et al. ClusterNet: Routing congestion prediction and optimization using netlist clustering and graph neural networks. *2023 IEEE/ACM International Conference on Computer Aided Design (ICCAD)*. IEEE, (2023).
13. Lyu, W. et al. An efficient Bayesian optimization approach for automated optimization of analog circuits. IEEE Transactions on Circuits and Systems *I: Regular Papers* 65.6: 1954-1967 (2017).
14. Chen, Z. et al. Artisan: Automated operational amplifier design via domain-specific large language model. *Proceedings of the 61st ACM/IEEE Design Automation Conference*. (2024).

4. AI for space information

4.1 Background

With the rapid development of 6G networks and low earth orbit (LEO) satellite constellations, communication systems are advancing toward a new era of 'AI-native' and 'space-air-ground integrated communication networks'[1-3]. 6G redefines the underlying architecture with ultra-low latency, ultra-high bandwidth and massive connectivity, while the space-ground network integrates LEO satellites, high-altitude platforms and terrestrial base stations to form a global multi-dimensional communication infrastructure.

In this context, large models have emerged as core enabling technologies for communication systems, supporting complex functionalities such as wireless channel prediction, network resource scheduling, and cross-domain signal processing. However, traditional large models face challenges in communication systems, including imbalances between computational power and efficiency, insufficient dynamic adaptability, and constraints on privacy and cost. Edge intelligence addresses these issues by offloading model inference to network edges (e.g., base stations and vehicle terminals) and using lightweight compression techniques to reduce latency.

Meanwhile, space-ground collaborative computing optimizes distributed deployment through an architecture integrating on-orbit computing, inter-satellite networking, and ground collaboration. As two key directions — split learning in wireless communication systems and split inference in space-ground communication systems — address these challenges by segmenting and collaboratively inferring large models, offering solutions for autonomous driving, smart healthcare, and disaster response[4].

4.2 Recent advances

4.2.1 Split learning for wireless communication systems

As 6G evolves toward AI-native networks, split learning has become a frontier for intelligent wireless communication[5-7]. The core idea involves dynamically partitioning a global model into lightweight client-side and server-side sub-models[7]. By integrating gradient aggregation and joint resource optimization, this approach overcomes the computational limitations of edge device.

Recent studies demonstrate that dynamic gradient aggregation reduces computational and communication loads by filtering critical features. And it eliminates the necessity of model exchange for model aggregation. Joint resource optimization strategies address device heterogeneity by employing intelligent sub-channel allocation, dynamic power control and model segmentation decisions, all of which contribute to reduced training latency and improved resource utilization.

©Yuichiro Chino / Moment / Getty

4.2.2 Split inference for space-ground communication systems

With the large-scale deployment of LEO satellite constellations, space-ground split inference has shown potential in applications such as disaster early warning, military situational awareness and climate modeling[8]. This technique splits large models into functional layers and distributes them between satellites and ground stations to optimize computational and transmission efficiency.

Modular segmentation based on transformer architectures can divide models into shallow sub-models deployed on satellites and deep sub-models on the ground[9-11]. Dynamic task scheduling adjusts strategies based on orbital positions and link quality, enhancing overall system performance.

Space-ground collaborative communication protocols incorporate techniques such as feature caching, pseudo-synchronous updates, sparse encoding, and quantization compression to reduce data transmission while maintaining accuracy. Furthermore, toolchains supporting open-source frameworks like Llama 3 and DeepSeek R1 integrate bandwidth simulators and energy consumption evaluation modules, accelerating the transition from simulation to deployment.

4.3 Key challenges and paths

4.3.1 Computational-communication coupling and heterogeneity challenges

Large-scale distributed computing tasks create exponentially increasing computational loads, while device heterogeneity introduces significant challenges in optimizing latency-energy balance during partitioning.

To address this, dynamic adaptive partitioning strategies should be developed, leveraging multi-agent deep reinforcement learning. These strategies, combined with inter-satellite networking, will enhance overall computational capacity.

4.3.2 Dynamic environment adaptation

Fluctuations in satellite-terrestrial links and wireless environments impact the stability of real-time inference. To tackle this, efficient fault-tolerant communication protocols can be designed, utilizing feature caching and data compression techniques to ensure inference continuity.

4.3.3 Privacy-efficiency trade-off

While gradient aggregation and data compression lead to information loss, transmission of remote sensing data also poses privacy risks. This can be addressed by introducing differential privacy and federated learning to optimize aggregation, paired with lightweight encryption to enhance the balance between privacy and efficiency.

4.3.4 Complexity of multi-dimensional resource optimization

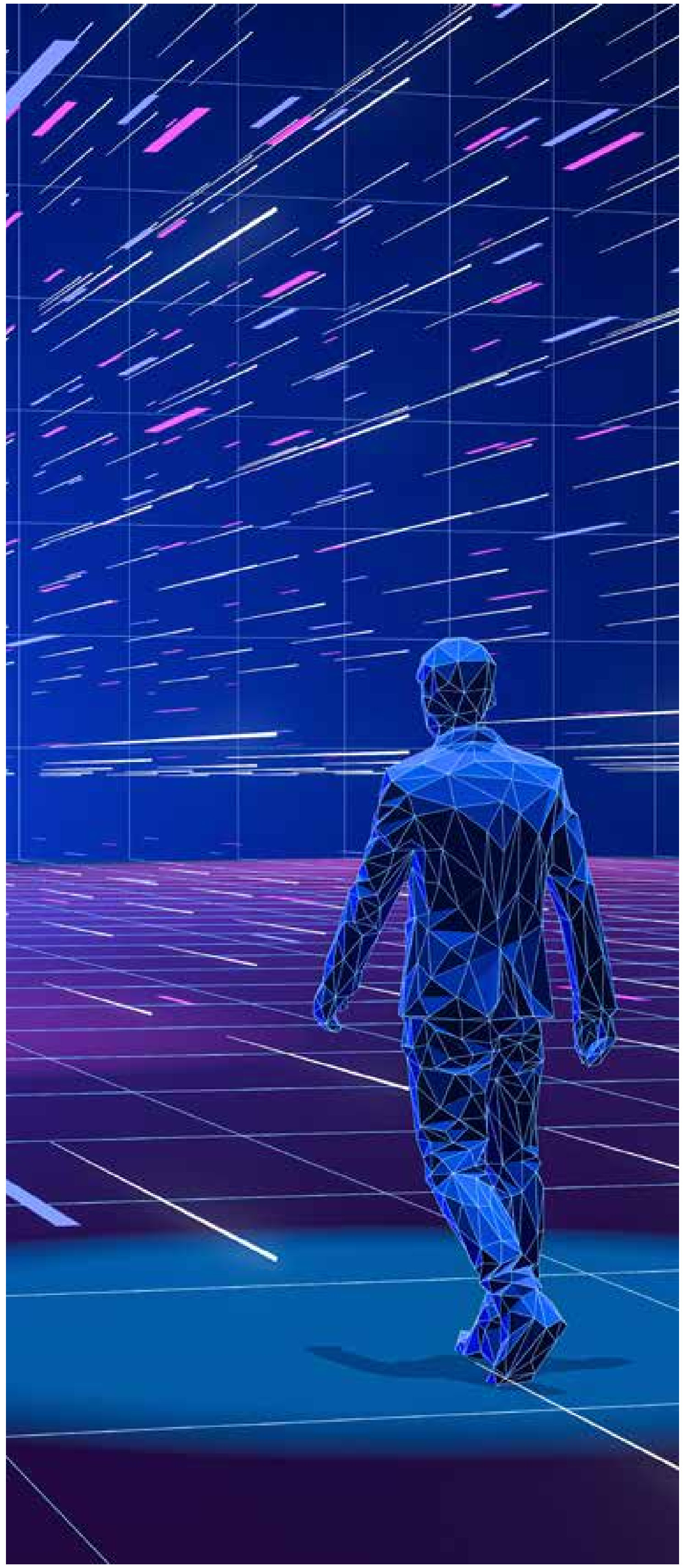
Joint optimization of channel allocation, power control and model partitioning forms a non-convex problem, making it challenging to meet millisecond-level latency demands. This can be resolved by establishing mathematical models for the convergence of split learning and split inference, quantifying the impact of resource constraints on performance to provide theoretical support for optimization.

References

1. Liu, L. et al. Democratizing direct-to-cell low earth orbit satellite networks. *GetMobile-Mob. Compu*. **28**, 5-10 (2024).
2. Li, Y. et al. A networking perspective on starlink's self-driving leo mega-constellation. *Proc. ACM MobiCom* (2023).
3. Li, L. et al. Opportunities and risks of satellite internet development in the context of 'new infrastructure'. *Satellite Application* **28**,38-42 (2020).
4. Abraham, et al. Classification and detection of natural disasters using machine learning and deep learning techniques: A review. *Earth Sci. Inform*. **17**, 869-891 (2024).
5. Lin, Z. et al. Fedsn: A federated learning framework over heterogeneous leo satellite networks. *IEEE T. Mobile Comput*. (2024).
6. Lin, Z. et al. Efficient parallel split learning over resource-constrained wireless edge networks. *IEEE T. Mobile Comput.* **23**, 9224-9239 (2024).
7. Lin, Z. et al. Splitlora: A split parameter-efficient fine-tuning framework for large language models. *arXiv preprint* **arXiv:2407.00952** (2024).
8. LEO Brings New Capabilities to Satellite Navigation. Available: https://www.salukitec.com/resource/leo-brings-new-capabilities-to-satellite-navigation/
9. Lin, Z. et al. Leo-split: A semi-supervised split learning framework over leo satellite networks. *arXiv preprint* **arXiv:2501.01293** (2025).
10. Zhang, Y. et al. Satfed: A resource-efficient leo satellite-assisted heterogeneous federated learning framework. *arXiv preprint* **arXiv:2409.13503** (2024).
11. Lin, Z. et al. Hierarchical split federated learning: Convergence analysis and system optimization. *arXiv preprint* **arXiv:2412.07197** (2024).

Chapter 8

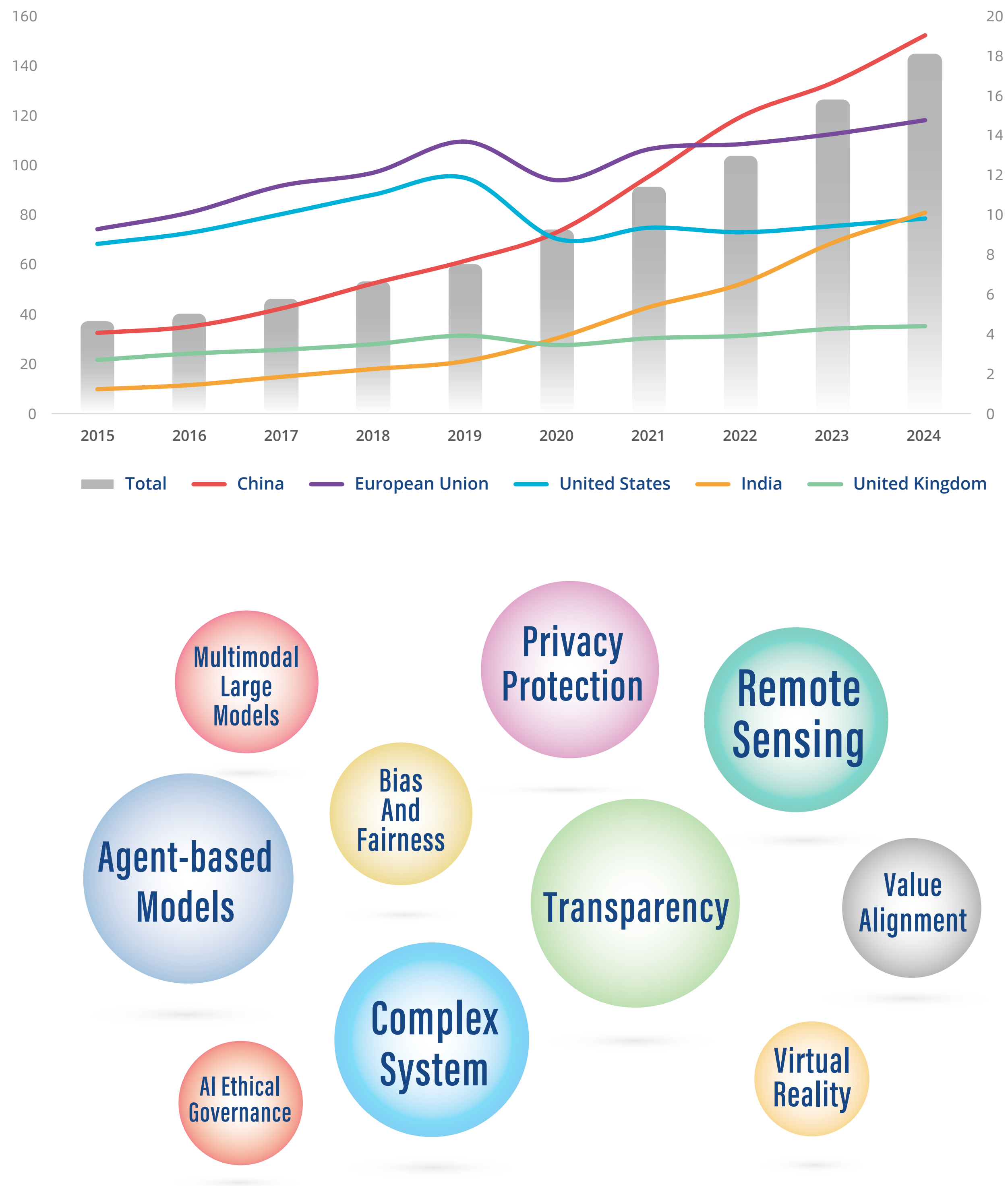
AI FOR HUMANITIES AND SOCIAL SCIENCES



©marian / Moment / Getty

Between 2015 and 2024, AI publications in humanities and social sciences nearly quadrupled, increasing from 37,100 to 144,800 (Figure 8). China surpassed the EU in 2022 to become the leading contributor in this field, while India rapidly caught up and overtook the US by 2024. Data analysis and the keyword cloud highlight transparency, ethics and governance as key areas of AI research, with a strong emphasis on mitigating societal risks through privacy protection, bias reduction and value alignment frameworks. From a methodological perspective, tools based on agent-based models, multimodal large models, and complex systems are reshaping how we analyse economic and social systems, deconstruct cultural phenomena, and advance new paradigms of human-AI interaction — offering profound implications for the sustainable development of the digital economy and society.

FIGURE 8 | Humanities and Social Sciences – Total AI Publications, National Trends (in thousands), and Keyword Cloud (2015–2024)



1. Computational social science

1.1 Background

The evolution of research paradigms in social sciences has historically followed four key stages: empirically driven qualitative research; theoretically driven quantitative research; mechanism-driven simulation modeling; and data-driven big data analysis.

AI technologies is now driving a fifth stage — one characterized by the dual engines of data and mechanisms. This new paradigm has the potential to profoundly reshape research methodologies in the humanities and social sciences, enhancing efficiency, scope, and depth of inquiry.

1.2 Recent advances

AI is catalyzing a threefold transformation in humanities and social science research: a shift in research subjects, a leap in model complexity, and a revolution in predictive capabilities.

1.2.1 Significant advances in data mining and processing

Generative AI has demonstrated strong performance in completing specific text classification tasks. Empirical studies show that generative models perform with high accuracy in identifying misinformation, classifying stance, and analysing sentiment without the need for large annotated datasets[1,2].

1.2.2 Enhanced modeling of complex systems

The development of agent-based modeling (ABM) has enabled significant breakthroughs in simulating complex social systems. ABM allows for the representation of bounded rationality among agents and captures non-equilibrium, non-transparent interaction patterns—greatly enriching the realism of simulations in socioeconomic systems. These characteristics make ABM particularly effective for analysing volatile, heterogeneous contexts such as political dynamics or financial markets.[3,4]

1.2.3 Upgraded capabilities in prediction and decision-making

Computational social science, which leverages large-scale data and high-throughput computing to model individual and collective behaviours, aims to support evidence-based decision-making.

At the individual level, the integration of AI algorithms with stochastic optimization has challenged the traditional 'predict-then-optimize' framework, offering improved performance[5,6] and adaptability in dynamic, complex operational environments[7].

At the policy level, ABM enables the simulation of multi-agent interactions under varying scenarios, helping to capture emergent societal outcomes and supporting the design of more responsive public policies.

1.3 Key challenges and paths

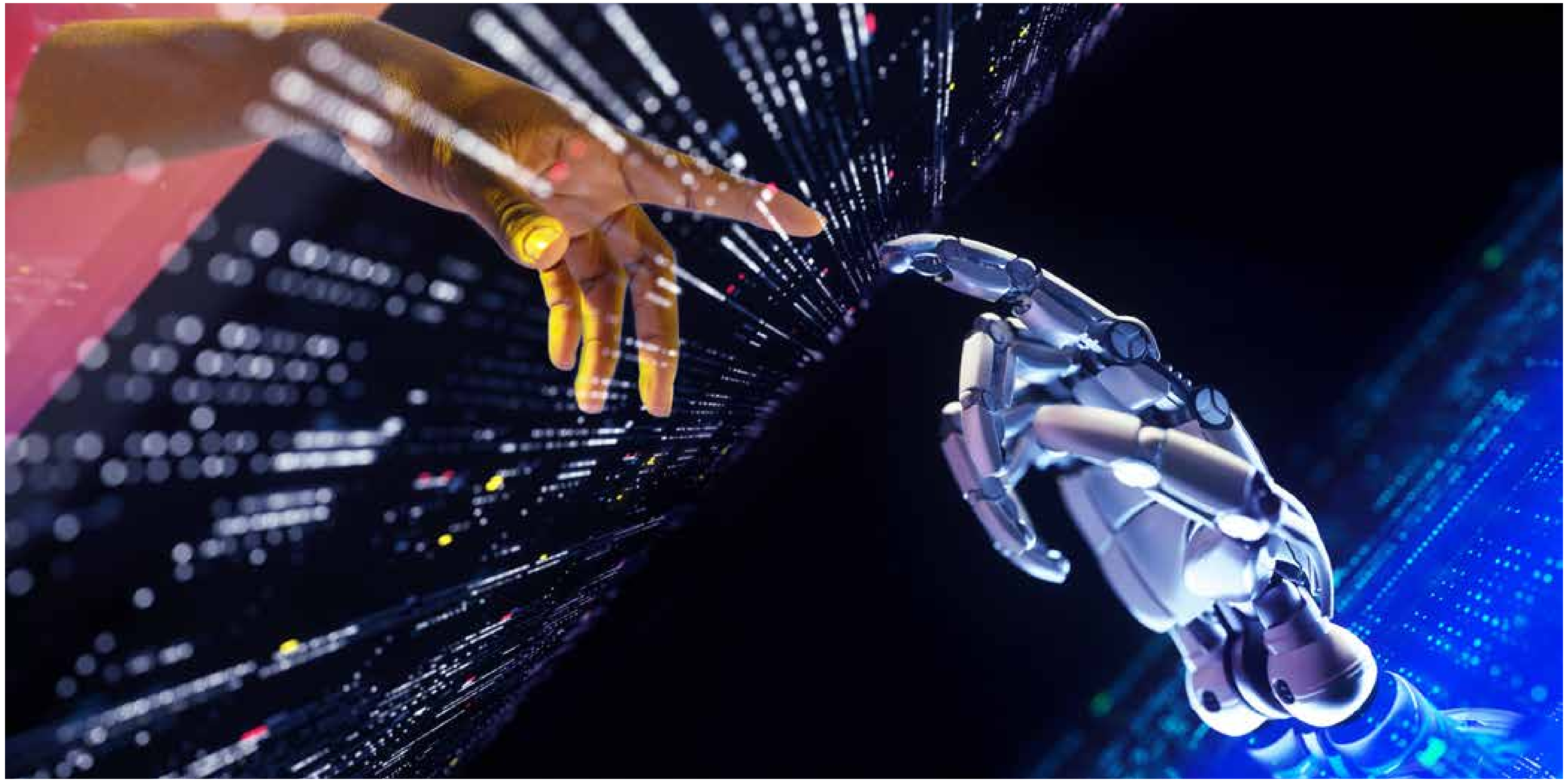
1.3.1 Key challenges and paths

Despite AI excels in prediction and analysis, its limitations in interpretability and mechanism integration pose challenges for theory-building — a core concern in the social sciences.

Future breakthroughs are likely focus on enhancing the synergy between large language models (LLMs) and ABM; fostering interdisciplinary collaboration to embed theoretical insights into algorithmic design; and adopting multi-scale modeling approaches to bridge macro- and micro-level analysis. These strategies will better reflect the dynamic evolution of complex social phenomena.

1.3.2 Uncertainty in mechanisms and the challenge of game-theoretic interactions

Human behavior, inherently nonlinear and stochastic, making it difficult to model with traditional approaches. This necessitates the incorporation of uncertainty quantification and game-theoretic reasoning into AI systems. Reinforcement learning and related adaptive techniques offer promising avenues for enabling AI to learn and simulate decision-making patterns among social agents—supporting dynamic forecasting and scenario-based policy simulation.



©Andriy Onufriyenko / Moment / Getty

1.3.3 Bias in large-scale models

While generative AI offers a powerful tool for social research, it also raises concerns about sample bias. Foundation models often reflect the dominant perspectives of populations with higher levels of education or income, those with more radical political views, or individuals residing in Western countries — while underrepresenting marginalized voices across political, religious and socioeconomic spectrums[8,9]. It would be shortsighted to assume that "silicon-based samples" can fully substitute for human data. Empirical scrutiny and continuous evaluation are essential to understanding both the potential and the limitations of these models. Moving forward, it is essential to adopt appropriate training methods, such as diversifying data sources and designing bias-mitigation algorithms, to progressively reduce biases in language models and advance generative AI toward greater fairness and inclusivity.

References

1. Ziems, C. et al. Can large language models transform computational social science? *Comput. Linguist.* **50**, 237-291(2024).
2. Krugmann, J. O. et al. Sentiment analysis in the age of generative AI. *Customer Needs and Solutions* **11**, 3(2024).
3. Schmitt, N. et al. Heterogeneous speculators and stock market dynamics: a simple agent-based computational model. *Eur. J. Finance* **0**,1-20(2020).
4. Gurgone, A. et al. Macroprudential capital buffers in heterogeneous banking networks: insights from an ABM with liquidity crises. *Eur. J. Finance* **0**, 1-47(2021).
5. Elmachtoub, A. N. et al, Decision trees for decision-making under the predict-then-optimize framework. *Proc. ICML* 2020, **268**, 2858-2867(2020).
6. Tulabandhula, T. et al. Machine Learning with Operational Costs, *J. Mach. Learn. Res.* **14**, 1989-2028(2013).
7. Qi, M. et al. A practical end-to-end inventory management model with deep learning. *Manage. Sci.* **69**,759-773(2023).
8. Durmus, E. et al. Towards measuring the representation of subjective global opinions in language models. *arXiv preprint* **arxiv:2305.17745** (2023).
9. Santurkar, S. et al. Whose opinions do language models reflect? *ICML*. **1244**, 29971-30004(2023).

2. Digital and intelligent humanities

2.1 Background

The integration of AI with technologies such as remote sensing, historical-geographic big data and virtual reality has introduced transformative approaches to uncovering, analysing and reconstructing historical realities. This convergence has driven significant innovation — from tomb detection and artifact identification, to cultural heritage conservation and the digital reconstruction of archaeological sites. It has deepened the analysis of spatiotemporal data and enabled multimodal reconstructions of the evolution of human civilizations, while accelerating the deciphering of ancient scripts and the study of semantic change.

2.2 Recent advances

2.2.1 Integration of AI and remote sensing for enhanced archaeological discovery

Traditional archaeological practices are often hampered by terrain and human activity, leading to site inaccessibility, uneven documentation, and subjectivity. The integration of AI with remote sensing technology offers new possibilities for large-scale, precise archaeological site detection. AI algorithms are capable of processing massive volumes of remote sensing data to automatically identify potential excavation sites[1,2]. This is particularly effective in multidimensional data mining, historical imagery analysis, and the recognition of complex geographic structures.

2.2.2 Multimodal AI for in-depth analysis of human and civilizational development

In the field of historical geography, natural language processing (NLP) techniques can be used to automatically extract key information from vast collections of historical texts. Meanwhile, computer vision technologies can analyse historical maps and images to identify geographic features and landmarks, generating detailed information layers. The integration of these diverse data sources contributes to a richer and more multidimensional understanding.

Using AI on ancient scripts and languages extends to tasks such as oracle bone inscription reconstruction[3], character recognition[4] and textual interpretation[5]. LLM-based agents can function as AI experts in paleography, supporting Chinese character education and the spreading of cultural knowledge.

2.2.3 Generative AI and the transformation of cultural display and dissemination

The integration of AI with virtual reality (VR) and augmented reality (AR) has introduced innovative ways to engage the public with cultural heritage. Applying advanced 3D modeling techniques in digital twin cities and metaverse environments, historical buildings and archaeological sites can be vividly reconstructed in digital space. Users can immerse themselves in historical scenes, experiencing ancient cultures firsthand and participating in virtual archaeological activities.

2.3. Key challenges and paths

The data used in the humanities is often heterogeneous and complex in origin, posing significant challenges for AI model training. The lack of standardized formats and variable data quality further complicate the modeling process. Standardization and quality control are urgent priorities. Researchers and technical experts must collaborate to develop unified protocols and rigorous evaluation standards to ensure the reliability and validity of datasets.

AI systems often encounter difficulties in interpreting complex historical and cultural contexts. One key challenge is perception. historical events and cultural phenomena are deeply embedded in specific social and historical contexts, and their meanings are often subtle and context-dependent. These intricacies are challenging for AI systems to fully grasp or accurately interpret. As such, human-AI collaboration is becoming increasingly important. AI should be seen as an auxiliary tool that enhances the efficiency of human scholars in analysis, interpretation and inference — rather than as a substitute for the depth of understanding and critical judgment that human experts provide.

Looking ahead, the future of digital humanities should focus on three strategic directions: human-AI collaboration, interdisciplinary integration, and technological innovation. Stronger collaboration between humanities scholars and AI practitioners is crucial for the development of intelligent tools that meet disciplinary needs while leveraging the strengths of AI. Cross-disciplinary database construction should be prioritized, integrating resources from history, literature, philosophy, sociology and related fields to build unified and high-quality research platforms. Finally, investment in talent development is critical — cultivating a new generation of scholars with both a solid foundation in the humanities and technical expertise in AI.

References

1. Berganzo-Besga, I. et al. Hybrid MSRM-based deep learning and multitemporal sentinel 2-based machine learning algorithm detects near 10k archaeological tumuli in Northwestern Iberia. *Remote Sensing* **13**, 4181(2021).
2. Mehrnoush, S. et al. Deep learning in archaeological remote sensing: automated qanat detection in the Kurdistan Region of Iraq. *Remote Sensing* **12**, 500(2020).
3. Zhang, C. et al. AI-powered oracle bone inscriptions recognition and fragments rejoining. Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence. International Joint Conferences on Artifical Intelligence, Yokohama (2020).
4. Guan, H. et al. Deciphering oracle bone language with diffusion models. the 62nd Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Bangkok, (2024).
5. 李春桃 , 等 . 基于深度学习技术的青铜鼎分期断代研究 . *出土文献* , **3**,16-32, 154-155(2023).

3. AI ethics and governance

3.1 Background

The impact of AI on humanity can be explored from short-, medium-, and long-term perspectives. In the short term, various intelligent tools have significantly enhanced work efficiency and convenience in everyday life. In the medium term, AI systems are transforming the ways society functions and reshaping underlying value structures. In the long term, the emergence of superintelligence may not only affect surface-level social activities but also raise fundamental questions of human existence.

With the rapid deployment of AI in sectors such as healthcare, autonomous driving and finance, risks are also emerging,[1,2] including concerns related to privacy breaches, algorithmic discrimination, divergent moral judgments, and accountability gaps in autonomous systems. Current research on AI ethics and governance largely focuses on emphasizing public policy, applied ethics, and the value that AI systems provide[3]. However, the potential implications of artificial general intelligence (AGI), strong AI, and particularly Artificial Superintelligence (ASI) are expected to become the central focus of future ethical and governance frameworks.

3.2 Recent advances

3.2.1 Transparency and explainability

Transparency and explainability remain central issues in AI ethics. Scholars have proposed a range of methods to improve the interpretability of AI systems. One widely adopted approach involves assigning importance scores to input features, thereby explaining the outputs of complex models[4]. This method has been applied extensively in domains such as finance and healthcare, helping stakeholders better understand the decision-making logic of AI systems.

3.2.2 Privacy protection

AI's capacity to process vast amounts of personal and biometric data has raised serious privacy concerns. In response, researchers have developed various protective strategies and technical solutions, such as federated learning and differential privacy, which aim to ensure data security while maintaining system performance[5,6].

3.2.3 Algorithmic bias

AI systems may replicate or even exacerbate existing societal biases, leading to unfair outcomes[7]. Some scholars advocate for the integration of ethical and legal principles into the design, training and deployment stages of AI models to mitigate these biases. Such efforts aim to ensure that the social benefits of AI technologies are realized without compromising fairness or equity[8].

3.2.4 Loss of control issues

With signs of AGI and even ASI increasingly becoming evident, superintelligent agents may lose control. These could cause scenarios like self-replication to evade human-imposed limits (e.g., shutdown problems) or unintended destructive outcomes driven by boundless and autonomous goal pursuits.

3.3 Key challenges and paths

3.3.1 Value alignment

Ensuring that AI systems align with human values and interests remains a key challenge in the ethics of AGI and strong AI. One promising approach involves training AI through human feedback to guide behaviour toward human expectations[9]. For example, the company. Anthropic, has proposed the '3H Principles' for AI alignment — helpful, honest and harmless — which have become influential in alignment research[10]. Looking forward, developing alignment frameworks that accommodate diverse cultural values across global contexts will be important.

3.3.2 Fairness and accountability

AI models can inherit biases from the datasets on which they are trained, leading to unjust decisions. Bias mitigation techniques aim to intervene at multiple stages of AI model development to reduce or eliminate such biases[11]. Fairness metrics have also emerged as essential tools for assessing and benchmarking AI system performance with respect to equity[12].

As AI increasingly participates in decision-making processes — such as in autonomous vehicles or automated medical diagnostics — clearly defining responsibility and accountability is crucial. Floridi and Cowls proposed five core principles of AI ethics: beneficence, non-maleficence, autonomy, justice and explicability. These principles underscore the fundamental importance of accountability in ethical AI systems.[13]

3.3.3 Loss of control and ethical response strategies

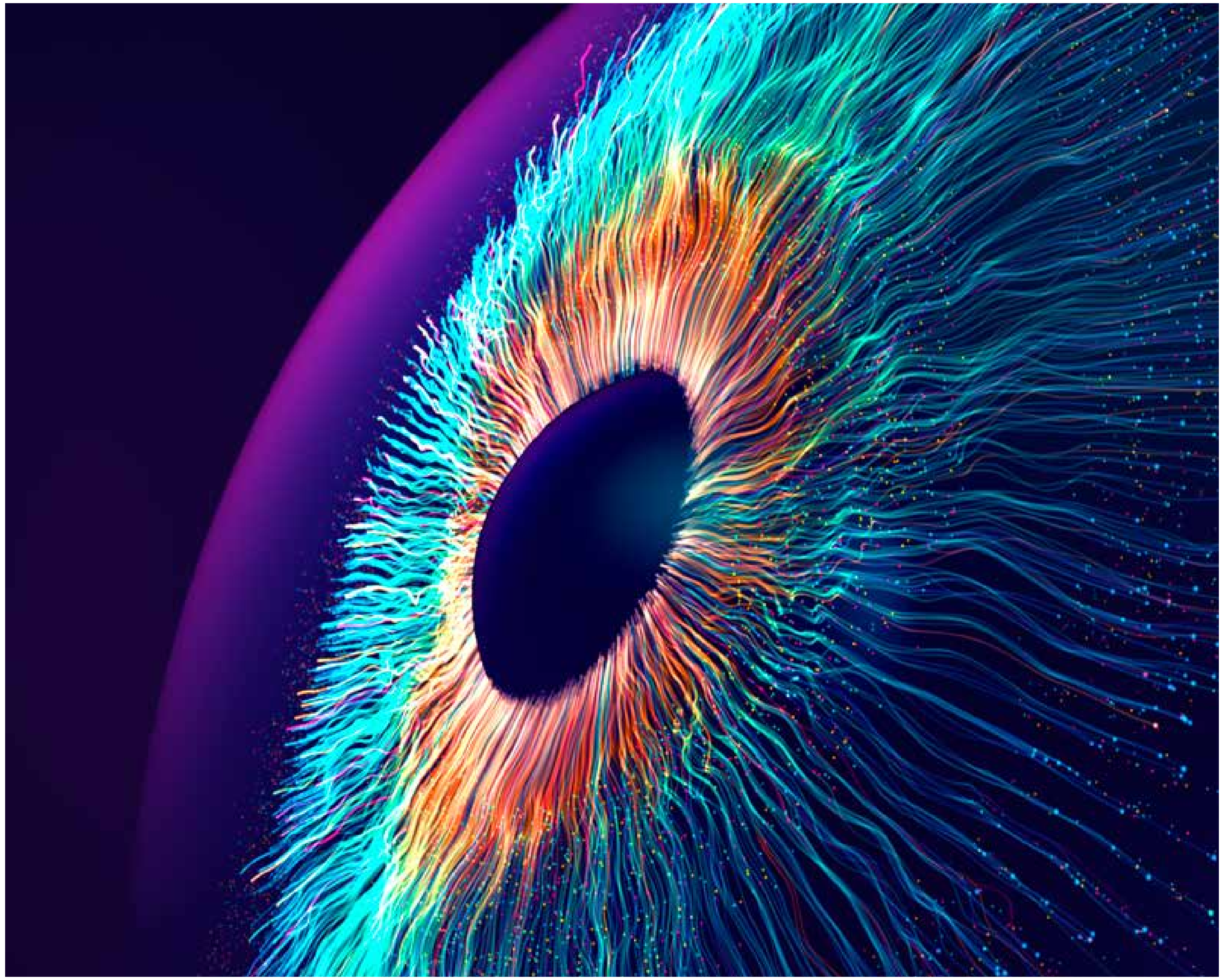
Ethical principles designed to manage superintelligence have evolved from human-based and human-centered to 'human-purpose-oriented' frameworks. Besides insights derived from Aristotelian virtue ethics[14], virtue dedication and intelligent contract ethics represent promising ethical response strategies.

References

1. Stilgoe, J. Machine learning, social learning and the governance of self-driving cars. *Soc. Stud. Sci.* **48**,25-56 (2018).
2. Verdiesen, I. et al. Integrating comprehensive human oversight in drone deployment: A conceptual framework applied to the case of military surveillance drones. *Information* **12**, (2021).
3. Birkstedt, T. et al. AI governance: themes, knowledge gaps and future agendas. *Internet Research* **33**,133-167(2023).
4. Lundberg, S. M. et al. A unified approach to interpreting model predictions. *NeurIPS* **30**,4765-4774 (2017).
5. McMahan, B. et al. Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics *(AISTATS)*. **54**,1273-1282 (2017).
6. Dwork, C. et al. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* **9**, 211-407 (2014).
7. Zhou, N. et al. Bias, Fairness and accountability with artificial intelligence and machine learning algorithms. *Int. Stat. Rev.* **90**,468-480 (2022).
8. Ntoutsi, E. et al. Bias in data-driven artificial intelligence systems-An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **10**, (2020).
9. Christiano, P. et al. Deep reinforcement learning from human preferences. *NeurIPS* **30**, 4299-4307(2017).
10. Bai, Y. et al. Constitutional AI: Harmlessness from AI feedback[EB/OL]. *arXiv preprint* **arXiv:2212.08073**, (2022).
11. Mehrabi, N, et al. A survey on bias and fairness in machine learning. *ACM CSUR* **54**, 1-35(2021).
12. Binns, R. Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*. 149-159 (2018).
13. Josiah Ober and Professor John Tasioulas. Lyceum Project - AI Ethics with Aristotle White Paper.https://www.oxford-aiethics.ox.ac.uk/lyceum-project-ai-ethics-aristotle-white-paper

Chapter 9

PROSPECTS AND POLICIES



©Andriy Onufriyenko / Moment / Getty

1. Future challenges and research directions

AI is rapidly penetrating major fields such as mathematics, materials science and life science — accelerating breakthroughs across more than 80 cutting-edge scientific problems. These challenges encompass foundational theoretical innovations in AI, paradigm shifts driven by interdisciplinary integration, and complex issues related to ethics and social governance.

1. 1 Evolution of AI models: from specialized to general-purpose

A core challenge is how to expand model capabilities to meet the diverse needs of real-world scenarios. A deeper understanding of the AI black box requires progress in the mathematical underpinnings of neural networks and the convergence mechanisms of reinforcement learning. The next generation of scaling laws and efficient reasoning methods will determine the performance limits of AI systems.

Looking ahead, general-purpose AI with cross-domain generalization capabilities may be achieved through multi-agent collaboration based on collective intelligence and unified modeling across all modalities. However, breakthroughs are still needed in adapting to dynamic environments and optimizing multi-dimensional resources.

1.2 Transdisciplinary integration: reshaping the paradigm of scientific research

AI is breaking down disciplinary boundaries, forming super-connectors across diverse scientific domains. For example, the integration of intelligent protein design and generative material models is enabling the creation of bio-responsive smart materials. In brain science, challenges in neural regulation are being tackled through brain-computer interfaces and bio-inspired models, driving the rapid development of next-generation neuromorphic computing devices.

AI-enabled multi-scale modeling frameworks are establishing unified theories that span from microscopic molecules to macroscopic ecosystems, uncovering common principles across disciplines. This transdisciplinary fusion initiates new hybrid fields such as 'bio-information-materials science', advancing research from analysis toward synthesis, and from fragmented understanding toward holistic insight.

1.3 Ethics and safety: building the 'brake system' for AI

As AI becomes increasingly embedded in critical domains like healthcare and embodied intelligence, safety and governance issues are becoming more pronounced than ever. At the data level, the tension between privacy protection and data sharing calls for innovations in new paradigms of encrypted computation. At the technical level, the lack of model interpretability and fairness leads to trust issues in applications such as clinical decision-making. At the value level, aligning AI with human values and constructing ethical frameworks will shape the direction of technological evolution.

There is an urgent need to establish dynamic risk assessment systems and ensure the safe and controllable development of AI through a synergistic combination of technological innovation and institutional design.

2. Policy framework

2.1 Policy objectives

With the rapid development of AI technology, the application potential of AI in research has become increasingly prominent. AI for Science (AI4S) aims to accelerate scientific innovation, enhance research efficiency and address complex challenges through the integration of AI technologies. Formulating comprehensive policies is critical to ensuring effective application of AI in research and to maximize its social and economic value.

The core objective of this policy is to accelerate the deep integration of AI technologies, promote interdisciplinary collaboration, and transform scientific discovery paradigms. Specifically, we aim to leverage AI to improve the efficiency of data analysis, optimize experimental design, and improve capabilities of simulation and prediction for problems that traditional research methods struggle to resolve. Additionally, the policy seeks to build an open, transparent and secure research ecosystem, encourage data sharing and knowledge exchange and reduce research barriers. To support the sustainable development of AI technologies, we will strengthen support for talent development, ethical governance and legal frameworks, ensuring that AI applications in research align with societal values and ethics, thereby advancing technological innovation and industry, and overall social well-being.

2.2 Principles and measures for policy-making

To achieve the strategic goals of AI4S, this policy will establish a systematic implementation plan across six core areas:

1. Data sharing:

data interoperability among research

institutions, governments and enterprises is encouraged. Establishing standardized scientific data-sharing platforms, and encourage open access and cross-disciplinary data integration. Developing data quality evaluation standards to ensure reliability and usability.

2. Security and privacy protection:
strengthening security management mechanisms for data, implementing strict access controls and encryption technologies. For data involving personal privacy and ethical sensitivities, establishing compliance review mechanisms to ensure alignment with ethical requirements and legal regulations in AI applications.

3. Algorithm development:
coordinating advances in foundational algorithm and domain-specific applications, building an open and collaborative innovation system. Promoting open-source sharing and iterative optimization of key algorithms, fostering industry-academia-research partnerships for technological breakthroughs.

4. Talent development:
Improving interdisciplinary education systems by integrating AI and scientific research. Encouraging universities to offer AI4S-related courses. Facilitating industry-academia collaborative training, launching specialized AI4S training programs, and cultivating interdisciplinary researchers with versatile competencies.

5. Funding support:
establishing dedicated research funds to support both fundamental and applied AI research. Developing multi-level funding mechanisms to provide long-term financial support for startups and research teams, accelerating the translation and implementation of innovations.

6. Legal and ethical governance:
establishing ethical guidelines for AI use, with defined accountability structures to ensure fair and equitable AI applications. Developing cross-departmental regulatory frameworks to ensure compliance with legal frameworks and foster responsible development under transparent oversight.

2.3 Mechanisms for policy implementation

1. Implementing bodies:
the adoption of AI4S policies involves the collaboration of multiple stakeholders. Government departments, such as the Ministry of Science and Technology and the Ministry of Finance, are responsible for policy-making and overall planning, jointly promoting relevant special deployments. Research institutions are tasked with technology development and application demonstration. National laboratories and universities conduct cutting-edge research by forming interdisciplinary teams. Enterprises, particularly technology companies, play an important role in the industrialization and commercialization of AI4S outcomes, contributing to the development of a robust innovation ecosystem.

2. Organizational structure:
constructing a multi-level, collaborative organizational framework. A national-level AI4S Strategic Committee should be established to coordinate interdepartmental resources and policies. At the regional level, AI4S promotion groups will be responsible for implementing policies and project docking. Meanwhile, alliances or communities among industry, academia, research and application should be encouraged to form an open cooperative innovation network.

3. Resource allocation:
integrating resources from multiple channels to ensure the development of AI4S. Financially, set up dedicated funds to support basic research and key technology development. In terms of research resources, it is important to promote sharing of computing power, data and models, such as through national artificial intelligence research resource sharing projects. Regarding talent resources, strengthen the cultivation of specialized AI4S professionals, establish related disciplines and training programmes.

2.4 Policy evaluation and adjustment mechanisms

1. Supervision mechanism:
a multi-dimensional supervision system should be established. Government departments are responsible for macro supervision of policy execution to ensure alignment with policy directions and objectives. Third-party evaluation agencies independently assess the effectiveness of project implementations, providing objective feedback. Meanwhile, introduce social supervision through mechanisms of transparent information release, accepting public oversight.

2. Feedback mechanism:
building dynamic feedback channels. Stakeholders should regularly report progress to supervisory bodies. Establishing expert advisory committees to gather opinions researchers and enterprise representatives, promptly feeding back issues and suggestions. Utilizing big data and artificial intelligence technologies to monitor real-time data on policy implementation, providing data support for feedback.

3. Evaluation and adjustment:
conducting regular, multi-dimensional evaluations of policy effectiveness, measuring impacts across technological progress, industrial driving force and social impact. Based on evaluation results and feedback information, policy content and resource allocation should be adjusted, continuously optimizing the implementation of policy.

Appendix I

Dimensions Database:
Dimensions is a research database which includes over 150 million publications, 42 million datasets, 7.9 million funding items, 167 million patent records, 0.9 million clinical trials records, and 2.5 million policy documents. By integrating and linking different types and levels of data, the Dimensions database effectively assists researchers in conducting multidimensional research. Researchers can analyze publications, citations, funding, and clinical trials at the levels of countries/regions, cities/metropolitan areas, institutions, and individuals. Researchers can also compare and analyze research topics, strengths in disciplines, and collaboration networks of different institutions or authors.

Nature Index:
The Nature Index provides simple, transparent and current metrics that demonstrate high-quality research and collaboration. The Nature Index database captures all affiliation information of primary research articles published within 145 natural-science and health-science journals that were selected based on reputation by a panel of active scientists, independently of Springer Nature.

Appendix II: Data Specification

1. AI publications
AI publications are based on publications (seed articles) in the AI research field from the Dimensions database. To cover a more comprehensive database, a model is constructed to iteratively match semantically similar publications using title and abstract embeddings of the seed articles. Additionally, the CCF-A classified proceedings, which are matched with the Dimension database, are included.

2. AI4S topic classification

2.1 According to requirements of the AI4S White paper, the AI publications are mapped to corresponding topics based on Field of Research (FOR) classifications:

2.2 Since the same publication can be classified into multiple FOR categories, if the publication belongs to the same AI4S topic, the number will be deduplicated under that topic.

3. Publications from unknown countries/regions are excluded when analyzing country/region-level data.

4. The illustration of scientific topics and AI methods under Core AI and AI4S topics (i.e., Figures 1.8 and word clouds in Figure 2-8) are illustrative diagrams. The relative sizes of the legends do not represent the actual data.

Notes:

1. Data updated: April 2025
2. Generative AI used for text polishing

FOR classifications	AI4S topic classifications
Information and Computing Sciences	Core AI
Earth Sciences	Earth and Environmental Sciences
Environmental Sciences	Earth and Environmental Sciences
Engineering	Engineering
Commerce, Management, Tourism and Services	Humanities and Social Sciences
Economics	Humanities and Social Sciences
Education	Humanities and Social Sciences
History, Heritage and Archaeology	Humanities and Social Sciences
Human Society	Humanities and Social Sciences
Language, Communication and Culture	Humanities and Social Sciences
Law and Legal Studies	Humanities and Social Sciences
Psychology	Humanities and Social Sciences
Biological Sciences	Life Sciences
Biomedical and Clinical Sciences	Life Sciences
Health Sciences	Life Sciences
Mathematical Sciences	Mathematics
Chemical Sciences	Physical Sciences
Physical Sciences	Physical Sciences

Nature Research Intelligence

Nature Research Intelligence (NRI) is committed to providing in-depth scientific research analysis and a panoramic overview of scientific research for policy makers, scientific research managers, research institutions, funding agencies and enterprises. By integrating and utilizing data sources such as Dimensions, Nature Index, Crossref, OpenAlex, etc., we correlate and analyze data such as scientific research articles, funding, patents, clinical trials and policy documents, and flexibly combine Nature Index, Nature Navigator, and Nature Strategy Report to provide objective and comprehensive scientific research performance data analysis, gain insights into scientific research development trends, reveal research potential and opportunities, and support strategic decision-making.

復旦大學

SAIS 上海科学智能研究院
Shanghai Academy of AI for Science

nature
research intelligence

©MF3d / E+ / Getty