科学智能白皮书

2025

復旦大學 SAIS

nature research intelligence

专家委员会

主席

金　力　　复旦大学

委员（按姓氏拼音字母排序）

步文博　复旦大学
龚新高　复旦大学
金亚秋　复旦大学
彭慧胜　复旦大学
漆　远　复旦大学、上海科学智能研究院
沈维孝　复旦大学
吴力波　复旦大学、上海科学智能研究院
张人禾　复旦大学

研究团队

第一章
徐增林　复旦大学、上海科学智能研究院
程　远　复旦大学、上海科学智能研究院
杨燕青　上海科学智能研究院
徐　燕　上海科学智能研究院

第二章
邱锡鹏　复旦大学
付彦伟　复旦大学
王守岩　复旦大学
杨　珉　复旦大学
邹　宏　复旦大学

第三章
陆　帅　复旦大学
石　磊　复旦大学
魏　轲　复旦大学
朱雪宁　复旦大学
高卫国　复旦大学
李颖洲　复旦大学
林　伟　复旦大学
杨　伟　复旦大学

第四章
向红军　复旦大学
季敏标　复旦大学
刘智攀　复旦大学
曹风雷　上海科学智能研究院
高　悦　复旦大学

第五章
应天雷　复旦大学
郁金泰　复旦大学
刘　雷　复旦大学
程　远　复旦大学、上海科学智能研究院
朱思语　复旦大学、上海科学智能研究院
彭汉川　复旦大学
徐书华　复旦大学

第六章
李　昊　复旦大学、上海科学智能研究院
张宏亮　复旦大学
赵　斌　复旦大学

第七章
迟　楠　复旦大学
徐　丰　复旦大学
刘　琦　复旦大学
曾　璇　复旦大学
杨　帆　复旦大学
高　跃　复旦大学

第八章
吴力波　复旦大学、上海科学智能研究院
唐世平　复旦大学
胡安宁　复旦大学
周葆华　复旦大学
吴肖乐　复旦大学
傅晓明　复旦大学
文少卿　复旦大学
杨庆峰　复旦大学
汤维祺　复旦大学

第九章
应天雷　复旦大学
颜　波　复旦大学

内容支持

吴文婷　施普林格・自然
张嘉慧　施普林格・自然
Rebecca Dargie　施普林格・自然
John Pickrell　施普林格・自然

数据支持

巨　蓉　施普林格・自然
黄珏珺　施普林格・自然
陈佳怡　施普林格・自然
Vivek Aggarwal　施普林格・自然

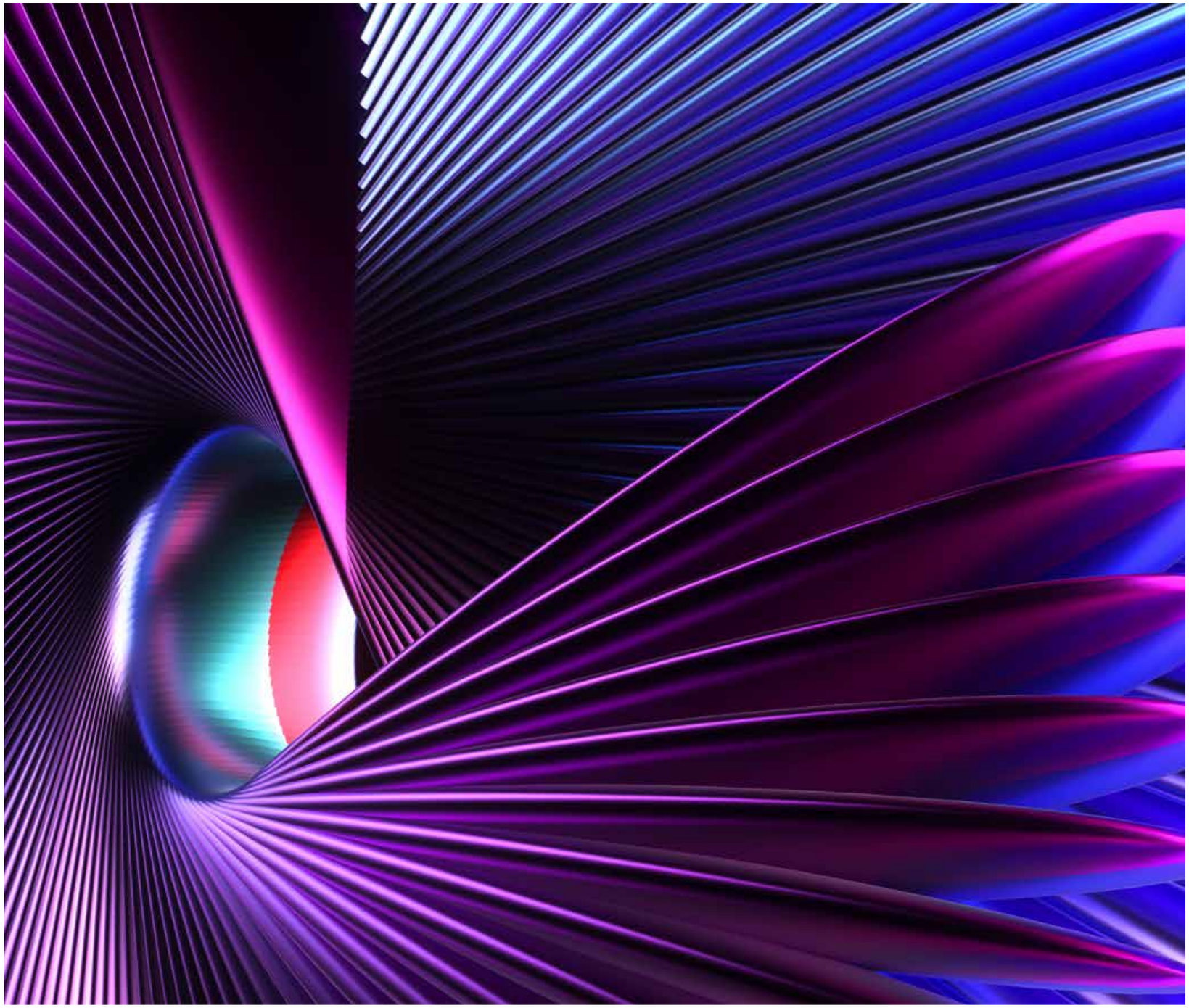
项目协调

徐晓创　复旦大学
杨燕青　上海科学智能研究院
王晓夏　施普林格・自然
丁思嘉　施普林格・自然
张瑶瑶　施普林格・自然

排版设计

赵新武　施普林格・自然
Sou Nakamura　施普林格・自然

目录



第一章

序言

1. 定义与范式

1.1 定义

面向科学研究的人工智能（AI）创新和人工智能驱动的科学研究的总和可被定义为科学智能（AI for Science, AI4S），是体现了人工智能创新与科学研究双向促进与深度融合[1]，从而变革科研范式。

1.2 范式

科学研究促进人工智能创新。传统科研范式大致可分为经验归纳（实验科学）、理论建模（理论科学）、计算模拟（计算科学）以及数据密集型科学[2]。实验科学由自然现象和实验结果归纳出一般性规律，但没有抽象出经验规律背后的普适理论。理论科学基于自然现象或实验结果，提炼科学问题并形成科学假设，然后运用逻辑推理和数学分析，构建普适理论，但难以在复杂系统中实验验证。计算科学以科学模型为基础，通过数值方法模拟复杂系统，但需要简化模型以及提高模拟精度，以解决模拟系统精度低且计算成本高的挑战。随着技术的发展和数据规模的增长，出现了数据密集型科学的研究范式。这一范式利用机器学习方法，自动从数据中发现统计关联，一定程度上避免了提出科学假设，但无法发现因果关系，且难以分析低质量数据和发现复杂系统中的规律。当前的科学研究主要面临系统复杂性的挑战，相互关联的自然、技术和人类系统受到跨时间和空间尺度作用力的影响，导致复杂的相互作用和涌现行为[1]。传统科学研究方法难以应对这些复杂性挑战，迫切需要新的科学研究方法。针对复杂数据中的因果关系，发展了一系列新的因果推断方法。针对高质量科学数据缺乏问题，如大气数据、天文数据等，发展了生成式人工智能技术，如扩散模型和大语言模型。针对处理复杂系统的局限性，发展了融合先验知识的深度学习，将先验知识嵌入深度神经网络，在增强模型可解释性的同时，显著提高模型的泛化能力，如物理信息神经网络[3]。

人工智能创新重塑传统科学研究过程，加速科学发现。人工智能通过融合数据和先验知识的模型驱动、假设生成与验证、自动与智能化实验以及跨学科合作等方式，加速科学发现。传统科学发现以实验观察和理论建模为核心，提出科学假设并归纳一般规律，如物理定律。人工智能则采用模型驱动的方式，从大规模数据中自动发现隐藏的规律，

传统科学发现从大规模解空间中生成候选假设并验证，效率低且难以找到高质量解[4]。人工智能凭借强大的数据处理和分析能力，可以更有效地探索解空间，生成高质量的候选假设。例如，在纯数学领域，机器学习可以辅助数学家发现新的猜想和定理[5]。科学研究依赖于实验评估理论。传统的实验设计和优化方法依赖人工经验和反复试错，成本高且效率低，如材料合成以及核聚变。人工智能与机器人技术结合可以实现实验的自动化设计与执行，并根据实时数据调整实验参数，优化实验流程和候选对象。

总之，人工智能可以有效整合不同学科的数据和知识，打破学科壁垒，促进多学科深度融合，解决学科的挑战性问题。跨学科合作不仅拓展了各学科的研究边界，还催生了计算生物学、量子机器学习、数字人文等新兴学科。

2. 发展与态势

2.1 最新进展

随着深度学习、生成模型与强化学习等技术的突破，人工智能不仅能从海量数据中识别人类难以察觉的复杂模式，更展现出自主提出科学假设、设计实验方案、优化研究路径的惊人能力。DeepMind 推出的 AlphaFold 3[1] 突破性地实现了对几乎所有分子类型的蛋白质结构预测，提高了蛋白 - 配体相互作用预测的准确度，为新药研发、疫苗设计带来革命性变革。Google 的 GraphCast 模型[2]、华为“盘古”大模型[3]、复旦大学“伏羲”大模型[4]等 AI 气象模型显著提升了全球天气预报能力，实现更长时间尺度、更高精度的天气预测。普林斯顿等离子物理实验室利用强化学习优化等离子体控制，解决撕裂不稳定性问题，加速核聚变能源的实现[5]。加州大学伯克利分校和劳伦斯伯克利国家实验室利用机器人执行实验，机器学习规划实验并结合主动学习优化实验过程，研发用于无机粉末固态合成的自动实验室 A-Lab，显著提高了材料合成效率[6]。

2.2 前沿科学问题与突破路径

2.2.1 如何构建跨尺度的科学智能模型

科学研究涉及从原子尺度到宏观系统的跨尺度建模，但当前 AI 模型通常仅适用于单一尺度，缺乏有效的多尺度耦合机制。

为了解决这一挑战，可以从以下几个方面寻找突破路径：

利用物理模型与 AI 的耦合建模，将已知的物理规律嵌入到 AI 模型中构建跨尺度关联，打造“灰盒模型”，提高模型的可信度和计算效率。开发跨尺度、多模态统一的神经网络架构，用于从微观到宏观的统一建模。

2.2.2 如何提升 AI 模型在科学研究中的泛化性

AI 模型依赖大规模训练数据，而高质量的科学数据往往有限。在数据有限的情况下，模型可能无法学习到有效的特征，难以适应新的领域或任务，限制了其在实际科学问题中的应用。

为了解决这一挑战，可以从以下几个方面寻找突破路径：

利用生成式模型生成高质量科学数据，补充数据稀缺领域的样本多样性。通过预训练跨领域基础模型，结合小样本学习技术，快速适应新任务或学科场景

2.2.3 如何利用 AI 拓展科学发现的创新边界

AI 目前仍局限于已有知识的重组与推理，主要通过对已有数据的模式识别和重组来生成结果，而缺乏真正的创造性思维。科学研究往往涉及跨学科的知识和数据，AI 模型在整合不同领域的知识时存在困难。如何使其真正参与科学假设的提出和验证，仍是未解的难题。

为了解决这一挑战，可以从以下几个方面寻找突破路径：

构建跨学科知识图谱、因果推理和生成模型，整合多领域知识库，使 AI 能够从已有知识中提取洞察并提出新颖的科学假设。建立强化学习驱动的 AI 辅助实验设计、数据分析、理论建模的闭环系统，实现自动化科学发现。开发可视化工具与交互界面，将 AI 生成的假设映射为可解释的科学逻辑链，支持领域专家进行修正与理论完善。

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).
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. 数据分析

本研究中，人工智能（AI）相关领域可以划分为: AI核心（如算法、机器学习等）、数学、物质科学、生命科学、地球与环境科学、工程科学、人文社会科学等七个领域。AI 核心以外的六个领域，统称为科学智能（AI4S）领域，后续章节将以上述领域划分展开。

根据自然科研智讯（Nature Research Intelligence）《自然》AI 相关出版物数量、引用量以及自然指数（Nature Index） 期刊等多源大数据，可以对 2015-2024 年全球 AI 相关出版物进行了系统性分析。数据结果显示，AI 和 AI4S 研究正经历规模扩张与范式变革的双重突破。

3.1 全球 AI 出版物迅猛成长，AI4S 加速井喷

2015 至 2024 年间，全球 AI 核心和 AI4S 领域的学术出版物总量快速增长。AI4S 异军突起，2020 年后加速成长，有力推动了 AI 研究整体的井喷态势。如图 1.1 所示，全球 AI 论文数量在过去十年间激增近三倍——从 30.89 万篇增至 95.45 万篇，年均增长率为 13.7%。2020 年是一个重要加速点，前后相较，年均增长率从 10.9% 跃升至 16.0%。同期 AI 核心领域论文占比从 44.5% 降至 38.0%，而 AI4S 占比相应提升了 6.4 个百分点，这源于 AI4S 论文的快速增长，2020 前年均增长 10.5%，2020 后年则以 19.3% 的速度扩张。其中，工程科学和生命科学最为突出，年均增长率从 2020 年前的 8.8% 和 15.3%，分别升至 16.1% 和 28.9%。

2015 年至 2024 年间，全球 AI 出版物排名前五的国家 / 地区格局发生了转变（图 1.2）。中国增长势头尤为显著，出版物总量从 2015 年的 6.01 万篇上升至 2024 年的 27.39 万篇，占全球总量的 28.7%。2018 年，中国 AI 出版物总量超越欧盟，居全球首位，2022 年超越欧盟和美国的总和。印度也展现出明显的追赶态势， 2015 年出版物总量为 1.82 万篇（仅为美国的 1/3），2024 年提升至 8.51 万篇，几乎与美国（8.57 万篇）齐平。

图1.1 | AI 出版物总量趋势与领域构成（2015-2024, 单位：千篇）

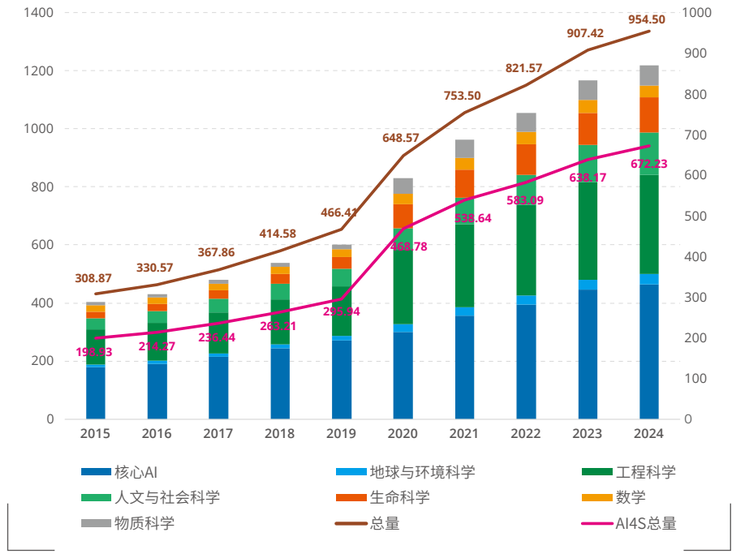


图1.2 | AI 出版物（前五国家/ 地区）总量趋势（2015-2024, 单位：千篇）

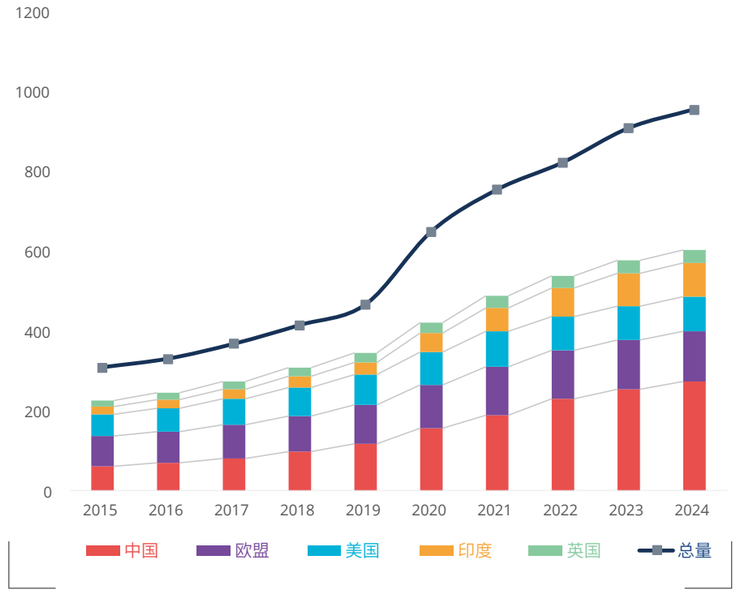


图1.3 | 自然指数AI 出版物引用总量（前五国家/ 地区）趋势（2015-2024，单位：千篇）

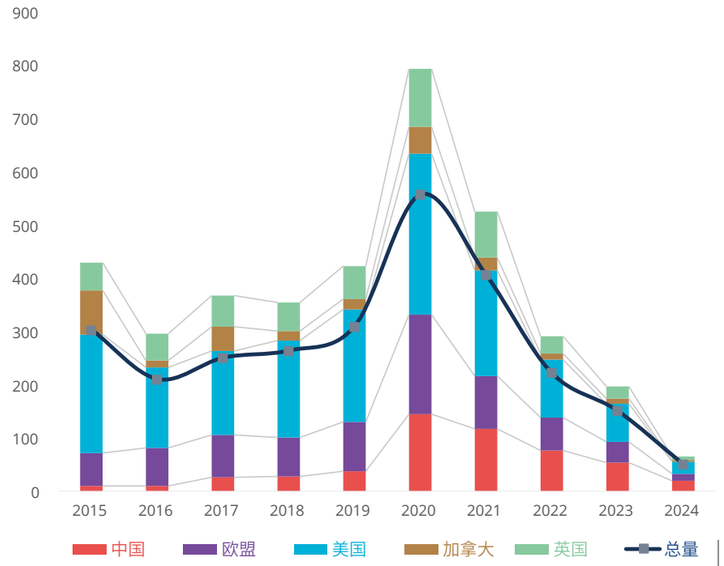
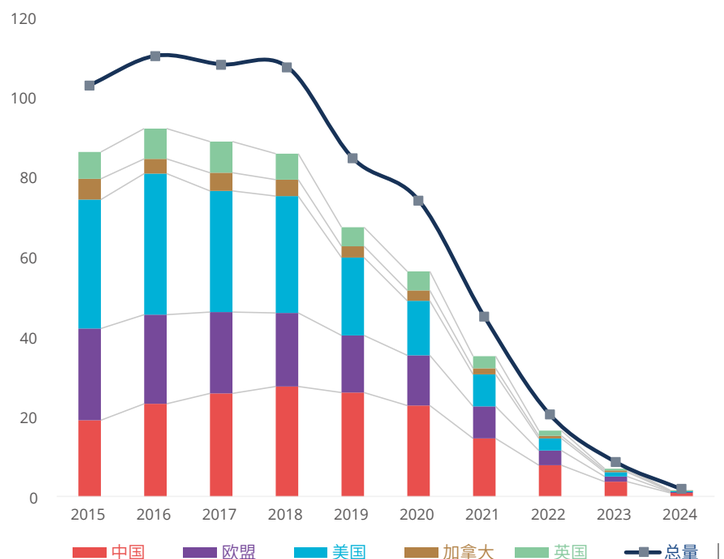


图1.4 | AI 出版物被专利、政策文档、临床试验的引用总量（前五国家/ 地区）（2015-2024，单位：千篇）



3.2 美国高质优势仍存，中国引领应用创新

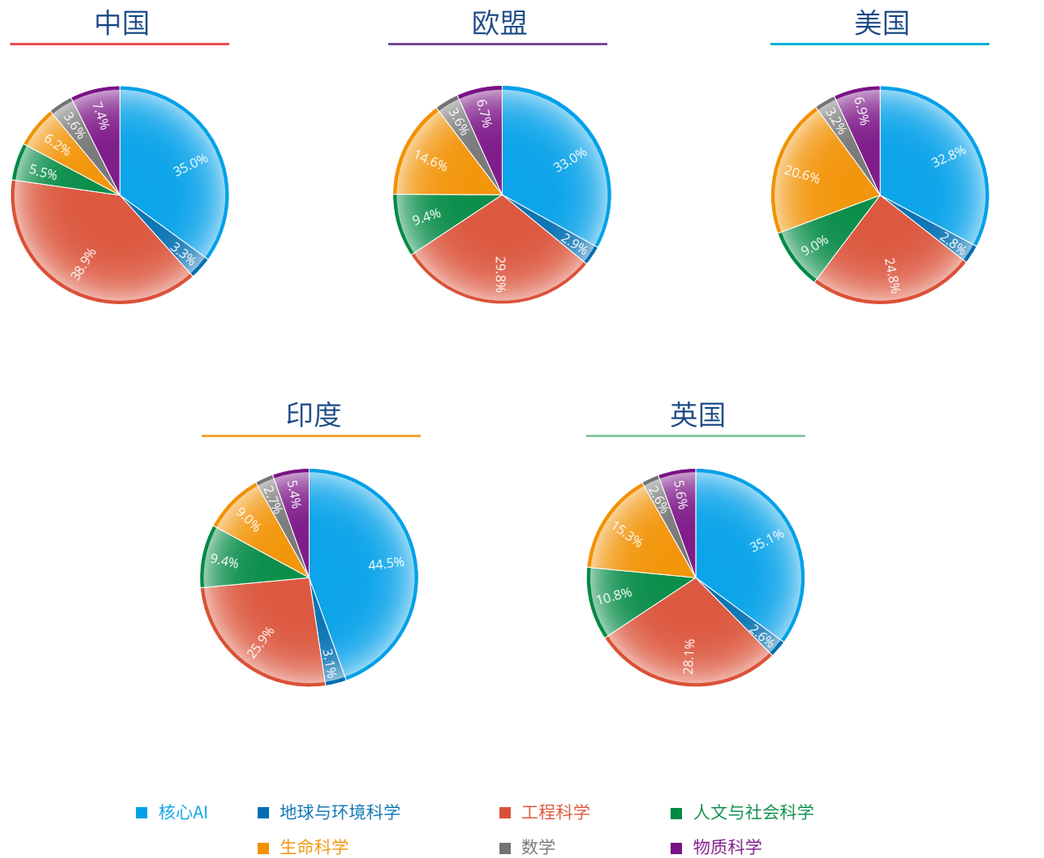
从科研质量上看，美国仍保持优势。如图 1.3 所示，基于自然指数追踪的高质量前沿研究期刊发表的 AI 相关论文引用量，美国始终保持领先地位，2020 年达到 30.28 万次。中国的崛起颇具颠覆性，引用量从 2015 年的 1.03 万次跃升至 2020 年的 14.48 万次，并于 2021 年首次超越欧盟。至 2024 年，中国的 AI 相关论文引用量占全球总量的 40.2%，实现了对美国（占全球总量的 42.9%）的快速追赶。需要指出的是，由于引用数据随着时间推移而积累，尽管目前统计显示 2020 年后引用数据趋势下降有所失真，但大体不影响国别间比较趋势分析。

中国在 AI 应用领域的创新同样体现了从“跟随者”到“引领者”的跨越。图 1.4 聚焦于专利、政策文档与临床试验中的引用数据。中国凭借持续高速增长，于 2016 年以 2.32 万次的引用量超越欧盟（2.23 万次），2019 年以 2.60 万次超越美国（1.96 万次）。至 2024 年，中国 AI 出版物在专利、政策文档与临床试验中的引用占比高达 41.6%，遥遥领先。同样，由于引用数据随着时间推移而积累，目前统计显示近年引用数据趋势下降有所失真，但大体不影响国别间比较趋势分析。

3.3 AI4S 国别优势各异，中美仍是最重要科研合作伙伴

2024 年 AI 出版物领域构成揭示了不同国家 / 地区在 AI4S 研究方向上的优势和特点（图 1.5）。美国、欧盟和英国聚焦工程科学（美 24.8%，欧盟 29.8%，英 28.1%）、生命科学（美 20.6%，欧盟 14.6%，英 15.3%）及人文社会科学（美 9.0%，欧盟 9.4%，英 10.8%）三大领域；中国以工程科学（38.9%）为主导，物质科学（7.4%）和生命科学（6.2%）次之；印度则形成工程科学（25.9%）为主，人文社会科学与生命科学（分别为 9.4% 和 9.0%）并行的研究格局。

图1.5 | AI 出版物总量（前五国家/地区）领域构成（2024，单位：%）



数据显示，AI 和 AI4S 的全球合作依旧稳步攀升。国际合作的 AI 出版物总量从 2015 年的 4.72 万篇跃升至 2024 年的 13.30 万篇，同期，国际合作的 AI4S 出版物总量从 2.99 万篇跃升至 9.48 万篇，均增长了三倍左右（图 1.6）。AI 出版物的中美合作在 2020 年到达顶峰后有所下滑，但依旧是全球规模最大的双边合作。至 2024 年，中美合作的 AI 出版物总量为 1.22 万篇，较 2015 年的 0.62 万篇增长近两倍（图 1.7）。同期，中国与欧盟、英国、加拿大及澳大利亚的科研纽带显著增强，中欧合作从 2015 年的 0.22 万篇增长至 2024 年的 0.98 万篇。

图1.6｜AI 和AI4S 出版物国际合作趋势（2015-2024，单位：篇）

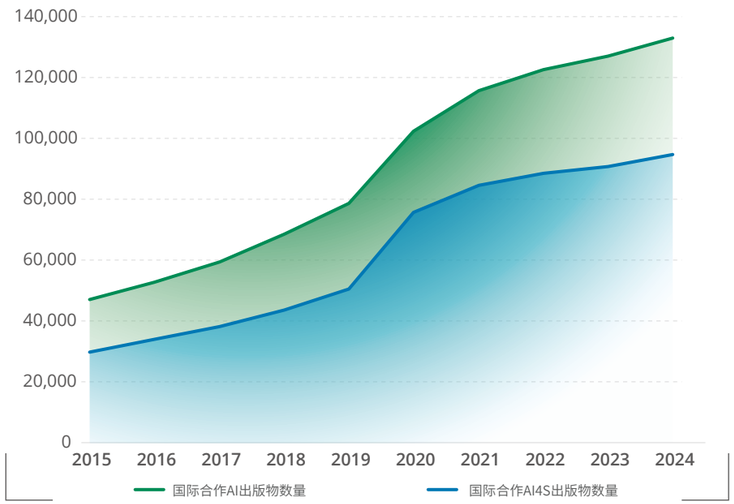
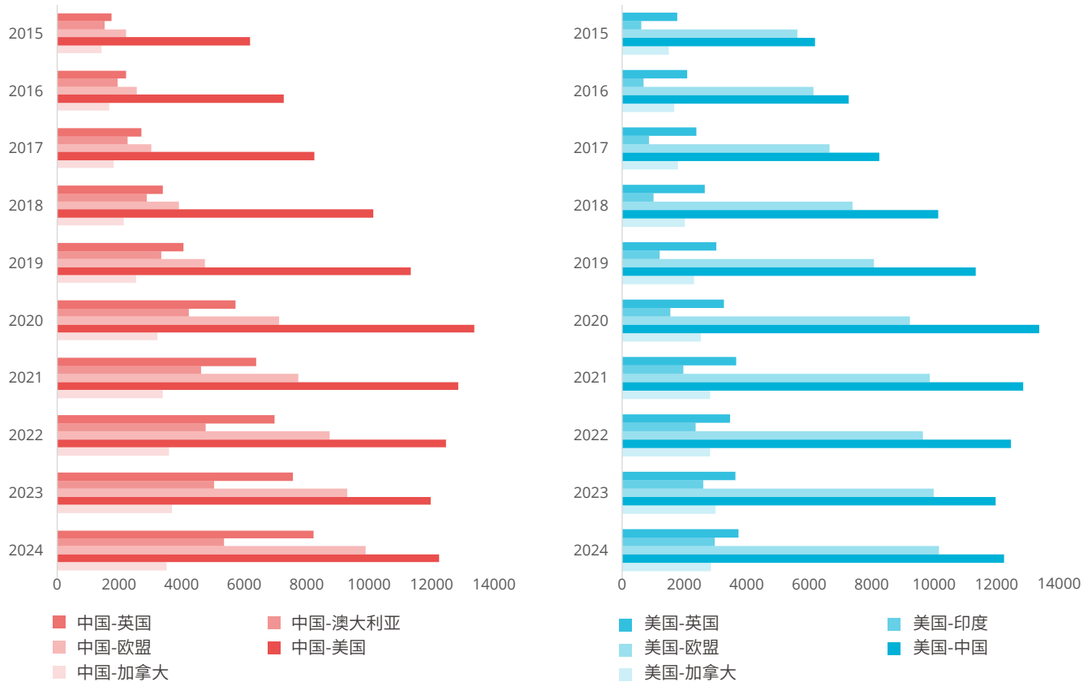


图1.7｜AI 出版物国际合作趋势- 中国vs 美国（前五国家/地区）（2015-2024，单位：篇）



3.4 AI4S 引领范式变革：哪些 AI 技术最获青睐？

2015 至 2024 年间，AI 催生了一场跨学科革命，其核心特征是领域科学和 AI 方法的深度融合与适配。通过科学家提出关键词和出版物数据库的匹配，可以发现在 AI4S 研究中运用最多的 AI 方法和技术（图 1.7）。如今，大语言模型（LLMs）已经成为物质科学、生命科学、社会科学等领域的通用科研工具。强化学习方法在工程系统控制、数学定理证明及物理模拟等复杂场景中占据主导地位。计算机视觉技术在生命科学和地球环境领域渗透显著。此外，分布式学习、图神经网络、可解释 AI 和边缘智能在不同学科中均得到广泛应用。AI 技术图谱揭示了一个根本性转变：AI 不仅是科学研究中可用工具集的扩展和创新，更是推动科学范式变革的“元技术”。一场 AI4S 革命，正在重塑人类科学发现的未来图景。

图1.8｜科学智能中最获青睐的AI技术（2015-2024）

第二章

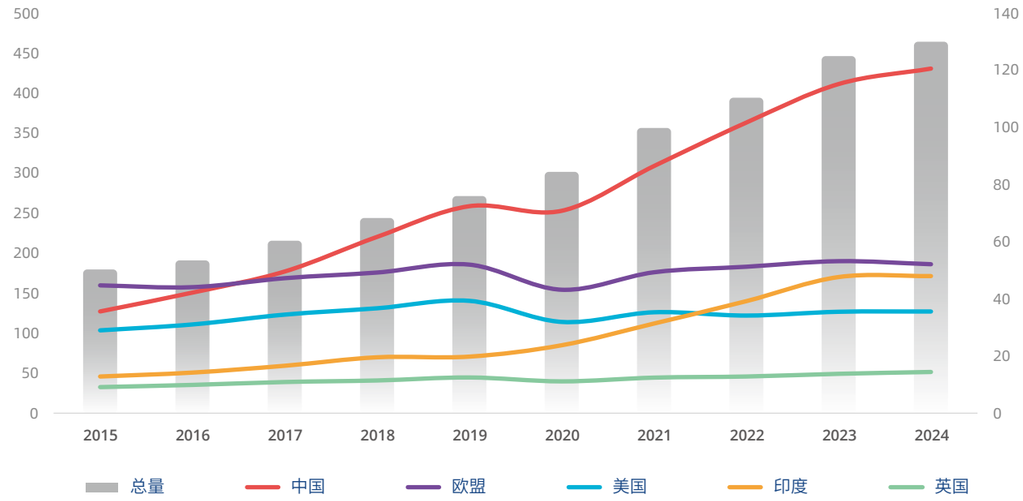
AI 前沿

©BlackJack3D / E+ / Getty

AI 核心

全球 AI 核心领域出版物从 2015 年的 17.92 万篇增长至 2024 年的 46.37 万篇（图 2），中国在总量方面优势显著，印度加速追赶，总量已超越美国，接近欧盟。从数据分析和关键词词云看，研究主题呈双聚焦：一方面，持续推进前沿模型、基础算法与计算架构创新；另一方面，2020 年之后，围绕 AI（内生）安全、对齐和极端风险的研究显著升温，展现出前沿模型技术创新和安全治理并重的发展趋势。

图2 | AI核心领域出版物总量、国家趋势（单位：千篇）与关键词词云（2015-2024）



1. 从大语言模型走向自主智能体

1.1 背景

近年来，以大语言模型（LLM）[1-3] 为代表的人工智能 (AI) 技术快速发展，已在多个领域展现出先前技术所不具备的涌现能力。借助千亿级参数规模，大语言模型不仅实现了海量知识整合和逻辑推理能力的突破，还展现出向通用人工智能演进的巨大潜力，具有重要的研究和应用价值。学术界与工业界纷纷将目光聚焦于大语言模型的相关研究，力图突破计算复杂度高、安全对齐难、可解释性弱等短板。与此同时，随着训练数据、算力资源逐渐面临增长瓶颈，研究者正积极寻找继续提升模型能力的第二扩展定律，推动模型向知识增强 [4]、多模态融合 [5,6] 和深度推理 [7,8] 方向演进，逐步催生出具备自主学习和决策能力的智能体系统 [9]。这一趋势不仅拓宽了人工智能的应用边界，更为实现真正意义上的通用人工智能 (AGI) 奠定了坚实基础。

1.2 最新进展

以大语言模型为核心的人工智能技术正迈入全新发展阶段。在训练数据和算力资源逐渐趋于饱和的背景下，研究者开始探索“第二扩展定律”，即从训练阶段的规模效应延伸到推理阶段，通过模型架构革新与软硬件协同设计，实现参数效率的指数级提升 [10]，进而大幅降低训练和推理的能耗。这一战略转变为持续推动模型能力升级提供了新路径。在技术演进方面，研究者聚焦于多个前沿方向：

1.2.1 知识增强：为弥补大语言模型长尾或领域知识不足、内在知识难以动态更新等问题，检索增强生成技术 [4] 通过对外部知识库的充分利用，使得模型能够快速获取专业知识和最新信息，有效提升了大语言模型的应用面和输出内容的可信度。

1.2.2 多模态融合：以 GPT-4o、Gemini 为代表的多模态大模型 [5,6] 通过跨模态对齐技术，实现了视觉、语音、文本等多模态信息的高效整合，大幅拓宽了模型应用场景。

3）深度推理：以 OpenAI o1/o3 和 DeepSeek R1[8] 为代表的推理模型，在解题回答中引入类似人类“思考—反思”的推理机制，用更长推理时间换取更高质量答案，在数学、科学和编程等复杂任务上取得显著突破。

4）自主智能体：多智能体系统 [9] 依托大语言模型的认知与推理能力，通过自主感知、任务规划、记忆系统及外部工具调用，显著提升了任务完成效率和系统协同能力。

5）安全可信：大语言模型安全可信研究正沿着 " 可解释性增强 - 价值对齐校准 - 可信评估体系 " 三位一体的技术路径纵深突破 [11]。

在应用层面，这些技术的突破正在引领人工智能迈向爆发性应用期。业界正积极将大语言模型及其衍生技术推广至办公助手、自动驾驶、智能教育、智慧医疗等多个领域，不断拓宽实际应用边界。当前的技术革新和应用实践预示着未来通用人工智能将不断向更高层次演进，为全球产业转型和社会发展带来深远影响。

1.3 前沿科学问题和突破路径

大语言模型需要在深度推理、扩展定律、高效架构、全模态模型、情感认知和群体智能等方向进行探索和突破，解决各自独特的前沿科学问题。

1）探索更高效且更通用的模型推理能力提升方法。优化强化学习策略与奖励信号设计，提高模型的学习与搜索效率，并利用人类反馈不断自我修正，突破复杂问题推理和长序列生成挑战，并将模型推理能力推广至更广阔的实际应用场景中。

2）寻找可以支撑模型能力提升的下一代扩展定律。在预训练和推理阶段的扩展定律之外，探索多智能体协作、物理世界交互和动态知识更新等下一代扩展方向。

3）设计软硬一体的高效模型架构。利用分布式训练、混合精度计算及专用硬件加速技术，加速模型训练和推理速度，使得大模型能够更快地响应实时任务需求，提升整体系统效率。

4）设计支持理解和生成的统一全模态模型。突破支持高效理解和生成的全模态模型架构，优化跨模态特征的融合与对齐，解决细粒度感知不足、幻觉频现以及空间能力欠缺等问题，为世界模型奠定基础。

5）探索大模型情感感知和认知调控技术。突破大模型的性格化和拟人化技术，增强大模型的跨模态情感感知能力，实现情境自适应的个性化情感调控机制，实现拟人化人机交互。

6）构建基于多智能体协作的自组织群体智能。探索复杂场景下的多智能体深度集成和自组织协作机制，构建可扩展的智能体协作框架，为未来人机共生智能化社会提供先行试验平台。

为更快地迈向通用人工智能，研究者需要突破上述大语言模型前沿科学问题，推动未来人工智能技术在科学研究、数字经济等场景中的落地与应用。

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. 具身智能

2.1 背景

具身智能是基于物理本体感知和行动的智能系统，通过与环境交互感知信息、规划任务、做出决策并控制本体完成任务，产生智能行为并持续演进。具身智能需可信、行为符合人类价值并能自我进化。

2.2 最新进展

具身智能技术近年来取得突破性进展，在感知、决策、控制及商业化应用等方面实现重大提升。高精度传感器增强了智能体的环境感知能力，融合感知与推理的动态四维世界模型提升了对复杂环境的理解与适应性，而强化学习与深度学习的结合则显著提高了自主决策与灵活应对水平。

在前沿研究方面，视觉 - 语言 - 动作（VLA）模型取得突破，谷歌 RT-2[1] 和 Physical Intelligence 的 π_0[2] 实现了从视觉与语言输入直接生成机器人动作的能力。Figure AI 推出的 Helix[3] 是全球首个人形机器人 VLA 模型，采用创新的双系统架构 [4]，为智能机器人控制提供了全新范式。人形机器人正加速演进，特斯拉 Optimus 二代 [5] 在工业与服务领域逐步落地，斯坦福大学的 Mobile ALOHA[6]、1X Technologies 公司的 NEO Gamma[7] 等多功能家务机器人展现出强大的实用能力，推动机器人在日常生活中的应用。

此外，具身智能在医疗、神经科学和虚拟现实等领域实现重要突破。复旦大学加福民团队开发脑脊接口技术，让瘫痪者重获行走能力；复旦大学类脑院构建千亿级神经元数字孪生平台 [8]，推动神经调控应用发展；VR 技术与具身智能结合，增强沉浸体验，推动未来人机交互革新。具身智能正迈向产业化，赋能医疗、制造、家用机器人及虚拟现实等多个领域。

2.3 前沿科学问题和突破路径

2.3.1 基础模型

现有具身智能基础模型的泛化能力有限，难以适应不同物体类别、场景和任务。如何提升跨本体、跨场景、跨任务的泛化能力，实现通用具身智能，是关键科学问题。

研发新一代具身基础模型，重点探索视觉 - 语言对齐（VLA）和双系统大模型。VLA 通过对齐视觉与语言信息，提升机器人对指令的理解与执行能力。双系统模型由反应系统和推理系统组成，分别负责即时响应与深度推理，以提升机器人决策能力。应用包括人机交互、机器人自主导航等，但在多模态学习与系统复杂度方面仍面临挑战，需要在效率与复杂度之间找到平衡。

2.3.2 数据引擎

具身智能系统的数据获取和融合仍面临数据质量参差不齐、数据标注成本高、跨模态数据同步性不足等问题，限制了模型的泛化和适应性。

研发多源异构数据采集与融合技术，构建高质量数据引擎。推动统一的数据采集与多模态融合平台建设，确保数据标准化与时间同步，促进开放数据集生态建设。多模态数据融合技术利用深度学习与融合算法整合视觉、触觉、听觉等信息，增强智能系统在复杂环境中的感知与决策能力，提高系统的适应性和可靠性。

2.3.3 交互能力

目前的具身智能交互方式仍显生硬，难以实现自然流畅的人机交互。如何提升具身智能的情感感知和交互能力，使其更自然、更人性化，是核心科学问题。

研究情感计算与多模态交互技术，优化具身智能体的互动体验。通过融合视觉、听觉和触觉信息，提高交互的沉浸感与效率。面对实时性与稳定性的挑战，需要确保系统能够快速响应，同时保持稳定数据传输。任务规划与执行层面，要求智能体具备自主探索与决策能力，优化数据获取、模型泛化和实时性。个性化与泛化性的平衡也是关键，需减少数据依赖，增强对不同用户和环境的适应能力。

2.3.4 本体研制

现有具身智能硬件在灵活性、感知精度和适应性方面仍存在局限，难以满足复杂任务的需求。如何打造兼具敏捷性与适应性的物理载体，是亟待解决的科学问题。

研发融合仿生结构、智能感知与先进驱动技术的新型具身智能硬件。利用多种传感器（视觉、触觉、听觉等）实现环境感知和自身状态监测。视觉感知依赖相机与 LiDAR 进行物体识别和深度感知，触觉传感器提供力反馈与纹理感知，听觉模块通过语音识别处理声音信号。运动控制方面，采用路径规划、动力学模型和协调控制方法，确保高效任务执行。同时，柔性电子与新型聚合物材料可提升感知精度，并通过集成设计增强数据采集能力和设备性能。

2.3.5 可信机制

具身智能的决策透明性和安全性仍存在较大挑战，智能体行为是否符合人类价值观、如何防范恶意操控和数据隐私泄露，是核心问题。

构建完善的可信评估与增强体系，以保障具身智能的可靠性。重点研究风险感知与价值对齐技术，使智能体的行为符合伦理规范和社会价值观。针对医疗、通信、娱乐等领域的应用，需加强安全防护机制，防范数据隐私泄露、恶意操控和未经授权访问等风险。同时，法律与伦理层面需要完善法规框架，以应对数据所有权、责任追究等问题，防止技术滥用带来的伦理困境。

2.3.6 具身智能评估

当前的具身智能评估体系尚不完善，缺乏统一基准，难以全面衡量智能体的控制能力、任务规划能力和泛化能力。如何建立科学合理的评估体系，是关键挑战。

研发系统化的具身智能评估框架，以更全面地衡量智能体的能力。需解决多方面问题，包括：现有评估方法偏重单一模态，忽略跨模态融合效果；智能体的泛化能力评估尚不充分；持续学习评估需解决灾难性遗忘问题；模拟环境与真实世界的差距影响评估准确性；多机协同任务的同步问题仍待优化；高层次决策任务与低层次执行任务的综合评估体系亟待建立。未来需跨学科合作，构建更完善的评估框架，以促进具身智能技术的发展与落地应用。

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. 脑机接口

3.1 背景

脑机接口技术为脑科学研究提供了全新的因果研究范式，并为脑疾病治疗开辟了靶向干预新路径。此外，该技术的发展不仅将驱动人工智能技术突破生物智能解析瓶颈，还将为类脑智能与具身智能的理论演进提供新路径，同时通过搭建人 - 机 - 环境智能融合接口系统，为构建可持续的智能社会生态系统奠定技术基础。

通过建立大脑与外部设备的直接通信连接，脑机接口技术实现了神经活动的记录、解码与刺激功能。当前，脑机接口正深度整合神经科学与人工智能技术，其发展轨迹已从单向的神经信息解析、神经调控信息写入，加速向脑机双向交互及脑智融合方向演进。

3.2 最新进展

神经信息运动解码研究已取得显著进展[1]，科学家们正在探索基于神经活动的言语、情感及意识解码技术[2,3]。基于神经信息解码建立疾病相关生物标记，为抑郁症等精神疾病的药物研发与神经调控精准治疗提供了重要应用依据[4]。神经调控技术通过光、声、电、磁等物理手段及化学方式调节神经活动，其中经颅磁刺激加速疗法为抑郁症治疗提供了创新性解决方案，超声神经调控在阿尔茨海默症等疾病治疗领域展现出广阔应用前景，光遗传学技术已进入人体临床试验阶段[5]，动能神经调控则可能成为未来新型神经调控手段[6]。

通过整合神经信息解码技术与神经调控及神经反馈机制，脑机接口可以实现高精度的外部设备控制或脑功能精准调控[7]，不仅使瘫痪患者恢复运动功能成为可能，更为抑郁症等精神疾病治疗提供了全新路径，实现了瘫痪患者重新行走[8]、抑郁症治疗[9]突破。美敦力公司研发的自适应闭环脑深部电刺激系统[10]已在欧美地区进入临床应用阶段，该系统通过实时监测与动态调控神经环路活动，显著提升了帕金森病等神经系统疾病的治疗效果。这一进展标志着脑机交互精准神经调控技术正式迈入临床应用新纪元。

当前人工智能技术的迭代升级正推动脑机接口研究迈向认知增强的新维度，其中“生物脑 - 智能体双向交互”已成为突破性方向。科学家融合脑机接口技术成功构建出具备自适应学习能力的虚拟大鼠智能体，发现虚拟控制网络中的激活状态准确预测了真实老鼠大脑中测得的神经活动，虚拟大鼠智能体能够模仿完成真实大鼠的所有复杂任务，甚至可以完成新的任务[11]。这项突破不仅为解析运动控制的神经机制提供了全新研究范式，更预示着“脑智融合科学”这一交叉学科的诞生，其在智能机器人、自学习神经调控系统、脑启发智能系统构建等方面具有重要应用前景。

脑机接口历经五十年发展，实现了从单向神经信号解析到双向信息交互脑机接口范式转换，并向脑功能与智能融合方向持续突破。多模态神经调控与解码算法融合将构建感知 - 决策 - 调控闭环系统，神经计算与 AI 融合推动脑机接口技术向认知增强技术发展，而交互维度将跨越神经信号解析至认知交互，物理接口扩展至虚拟现实、智能环境等多模态协同。神经机制、神经技术与人工智能深度融合的发展趋势预示着脑机接口技术正式迈入智能增强与认知重塑的新阶段。

3.3 前沿科学问题和突破路径

人脑作为大规模的复杂动力系统，其网络化连接模式、动态时变特性和非线性交互机制，构成了神经解码与编码技术实现可靠性和稳定性的核心挑战。

如何实现神经集群复杂功能的特异性调控？针对兴奋性 / 抑制性神经元、感觉 / 运动神经元等特定神经群体，通过跨时空尺度的神经信息解码与编码研究，结合超声、光学等技术手段建立从单神经元到神经环路层级的毫秒级时间精度与微米级空间分辨率的神经功能监测调控体系，将为神经疾病与精神障碍治疗提供特异性神经调控解决方案。

如何实现神经元物理与化学交互作用的精准调控？脑功能实现机制不仅涉及神经元电活动及神经环路动态互作，更依赖于分子层面的调控要素（包括神经递质、受体蛋白、离子通道等）。通过整合分子调控网络与神经刺激技术体系，可在多维度实现神经系统的分层精准调控，进而达成更高效、更精细的全脑神经功能干预目标。

如何实现脑功能动态过程的精准调控？建立基于神经生理与神经递质等信息的疾病特异性生物标记体系，构建神经环路动态因果计算模型，开发集成感知 - 存储 - 计算 - 控制功能的智能神经调控芯片，最终形成具备神经信息实时感知、动态建模与智能决策能力的闭环脑机交互系统，实现神经核团、神经环路功能动态自适应调控。

如何实现脑智融合自学习策略？将多模态神经信息编码框架与具身智能大模型进行深度融合，通过构建自然化人机交互学习范式，将显著增强智能体的自主决策能力与环境适应性。通过动态自学习机制实现神经信息解码与功能重编程，不仅能驱动实体机器人及智能设备的精准操控，还可建立虚拟世界中智能体间的多模态交互通道，最终形成具备沉浸式体验与实时神经与认知功能响应能力的脑智融合系统。

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. Sahel, JA. et al. Partial recovery of visual function in a blind patient after optogenetic therapy. *Nat. Med.* **27**, 1223-1229 (2021).
6. Yang, M. et al. Intermittent vibration induces sleep via an allatostatin A-GABA signaling pathway and provides broad benefits in Alzheimer's disease models. *Adv. Sci.* **12**, 2411768 (2025).
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 内生安全

4.1 背景

面对新一轮生成式人工智能浪潮，如何兼顾发展与安全已成为国际社会的共性难题[1]：一方面，AI 系统在数据采集、模型训练到实际部署推理全生命周期均存在安全漏洞，现有安全防护体系在应对新型威胁时往往出现系统性失效；另一方面，大模型智能体技术范式快速演进，其依托 AI 系统软件将基础大模型的生成内容转换为作用于数字或物理世界的行为，若缺少安全管控，将导致 AI 生成内容风险向物理域、社会域快速外溢。因此，将内生安全理念融入基础大模型研发、AI 系统软件设计和训练部署全过程，是构建自主可控的内生智能防护体系的关键[2]。

4.2 最新进展

AI 系统全生命周期均面临严峻安全挑战：数据采集阶段，投毒攻击通过注入噪声或恶意样本，误导模型学习；训练阶段，通过植入模型后门，使其具备隐蔽功能；推理阶段，对抗样本、模型幻觉和风险内容生成已成为主要威胁。超大规模参数、极深网络结构的基础大模型生成行为可解释性差、安全对齐内生性不强，易遭受越狱、提示词注入、隐蔽后门等对抗诱导[3]。当前多采用的“外部加固”策略难以适应 AI 系统的动态演进和模型复杂特性。当前 AI 安全领域尚未形成完备理论框架，缺乏对智能系统自身安全能力的度量方法。

另一方面，基础大模型行为所具有的非预设性，使得前沿 AI 系统或带来重大红线风险： 国内外已关注到前沿 AI 系统已突破自我复制[4]、欺骗[5]等红线能力边界。因此，需通过 AI 内生安全理论原始创新，建立内生治理与外部管控相结合的安全体系。

4.3 前沿科学问题和突破路径

4.3.1 AI 内生安全理论

AI 内生安全需在模型设计之初嵌入安全机制，实现“安全即特性”目标，例如：通过自适应机制自动调整模型参数；结合可信执行环境保护分布式训练隐私安全。构建 AI 系统内生安全架构，整合可信计算、零信任框架是重要研究方向。

突破路径：传统安全范式试图通过“封堵查杀”的被动防御手段干预外因，无法化解系统构造缺陷引发的本源性矛盾。动态异构冗余架构[6]将安全属性从代码的脆弱性转向架构的确定性，使内生安全矛盾在系统层实现演进转化或动态和解[2]。动态异构冗余架构通过多模型集成、异构算法模型、多态执行体与策略化调度机制的深度融合，使攻击者难以捕捉稳定的攻击界面；通过算法异构化部署、模型动态迁移及特征空间重构，打破攻击所需的静态环境假设，增强 AI 系统在复杂环境中的稳定性和安全性[7]。

4.3.2 AI 系统安全评估与防护

现有安全评估依赖静态测试和对抗攻击实验，难以全面评估开放环境下 AI 系统的安全性。为持续监测通用大模型的安全风险，发展自动化、覆盖面广、风险发现能力强的 AI 动态安全评测技术尤为重要。

突破路径：构建自动化安全评估工具，基于博弈论和机器学习开发自适应评估方法，探索通用安全标准；研究面向内生安全的大模型风险靶向挖掘技术，发现代表性风险用例，形成风险数据库；研究基础大模型内容风险控制机制，包括算法、内容规范，防治 AI 产生破坏性的虚假信息。

4.3.3 前沿 AI 系统风险感知与治理

面向前沿 AI 系统的红线风险，研制主动风险感知技术，建模智能体行为失控机制，形成系统化、实操性强的风险评估体系，对预警和治理前沿 AI 系统至关重要。

突破路径：研究工具交互、环境感知、思维推理、记忆增强等优化方法，动态激发基础大模型潜能，主动感知风险红线突破点；从大模型、训练数据和系统软件出发，研究面向大模型关键危险能力的抑制方法，实现面向红线的大模型行为编辑与对齐方法，构建具备智能风险感知能力的 AI 系统软件。

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. 吴锟等 . 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).

第三章

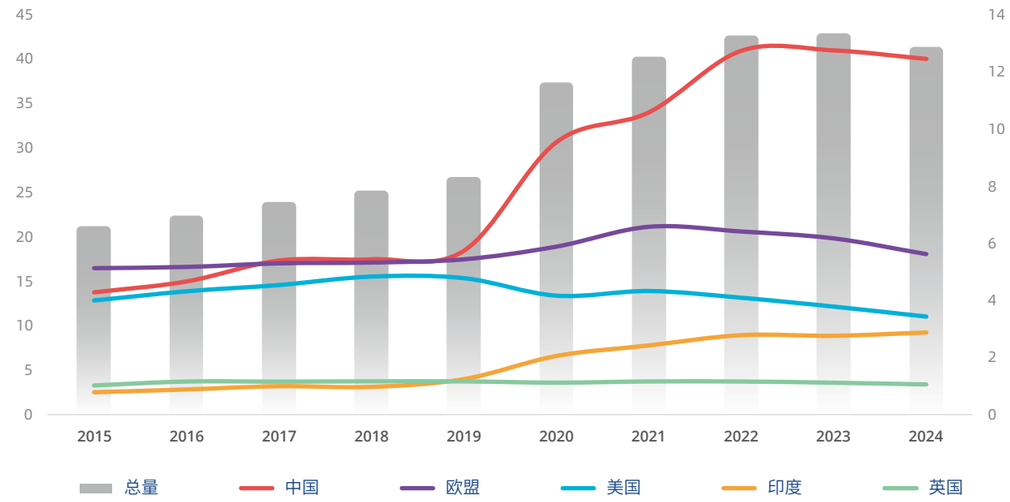
数学

©Constantine Johnny / Moment / Getty

AI 与数学

2015 至 2024 年间，全球数学领域 AI 出版物从 2.12 万篇增至 4.12 万篇，呈现近两倍增长（图 3）。中国于 2017 年后持续超越欧盟与美国；印度 2020 年后增长加速，逐步拉近与美国的差距。在基础理论、模型设计以及算法层面，数学是 AI 理论和创新之基础。在运筹优化、科学计算和复杂系统等领域，数学和 AI 深度融合，协同创新。数据分析和关键词词云显示，强化学习、循环神经网络、生成模型和扩散模型等关键词最被这一领域科学家看重，体现了底层理论创新和跨学科融合的范式演进。

图3｜数学领域AI出版物总量、国家趋势（单位：千篇）与关键词词云（2015-2024）



1. 基础理论

以数学为视角，利用数学理论与数值方法推动人工智能理论的发展，其研究可分为三大层面：人工智能基础理论、模型设计以及算法实现。

在人工智能（AI）基础理论方面，核心问题之一是分析深度学习模型的表达能力。数学工具如函数空间、逼近论和数值分析为揭示神经网络内在的非线性结构提供了方法论支持，并奠定了模型稳定性和泛化能力的理论基础。研究者借助这些工具构造出如巴伦空间、变分空间等由神经网络诱导的函数空间，从而量化网络表达能力。同时，通用近似定理证明通过加宽或者加深前馈神经网络能够以任意精度逼近紧集上任意连续函数。尤其是在逼近本质低维的高维函数时，神经网络可以有效回避高维维度灾难问题，这在一定程度上解释了深度学习处理复杂模式的能力。对于其他网络结构，数学同样发挥着作用：图神经网络（GNN）的信息传播和聚合过程可通过图论和图拉普拉斯算子的谱分析来解释；将深度学习模型视作动力系统，通过微分方程和稳定性理论分析循环神经网络（RNN）的隐藏状态演化，不仅揭示其长序列稳定性，也预示激活函数选择不当时可能引起的梯度消失风险；动力系统中的平衡点、吸引子和分岔理论进一步为神经网络训练过程中的动态行为提供理论支撑，指导更稳定高效的算法设计。

在人工智能模型设计方面，数学理论指导着网络结构设计、学习范式等。例如，扩散模型的架构根植于概率论和随机过程，其正向过程利用马尔可夫链逐步注入高斯噪声使数据退化，逆向过程则依赖参数化条件概率逐步剥离噪声实现重构，构成一个可逆随机过程。这样的设计在确保生成样本多样性的同时，也保证了高质量的重构。此外，RNN 利用递归结构捕捉时序数据的动态特性，其状态更新可用差分或微分方程描述，而残差网络（ResNet）则通过引入跨层“捷径”联结，缓解深层网络中梯度消失问题，其理论分析依赖于线性代数和微分方程，从而解释了信息恒等传递的原理。

在人工智能算法实现层面，数学不仅体现在数据预处理、损失函数构造上，也贯穿于优化算法设计。数据预处理方面，通过统计建模与样本生成技术改善数据质量：针对缺失值问题，可依据数据分布（如高斯假设）进行参数估计并采用概率插补；面对类别不平衡，可利用过采样或线性插值在特征空间内扩充少数类样本，从而增强模型鲁棒性。损失函数设计中，正则化方法（如 L1 正则化促进稀疏性、L2 正则化限制参数幅度）起到控制模型复杂度和防止过拟合的作用，此外还可依据神经网络诱导的新型函数空间设计专门的正则项。优化算法和数值算法分析皆依赖于数学理论，不仅有助于训练过程的高效实现，还可以为模型部署提供支持，例如利用张量分解降低计算复杂度，或通过拓扑分析优化硬件适配。

本部分的前沿科学问题可以分为两类，一类聚焦于揭示人工智能模型结构内蕴的数学理论；另一类着眼于人工智能算法分析。第一类前沿科学问题通过先分析一般简单深度神经网络模型的表达能力，再分析复杂常用模型的表达能力，为模型的可解释性奠定数学基础，最后基于理论分析指导模型设计。第二类中的关键前沿科学问题之一是泛化性。正则化、隐式正则化理论研究已经为模型泛化性奠定了基础，进一步将求解问题的数学理论融入算法设计，有望得到具有一定泛化能力的训练模型。

总之，数学为深度学习和人工智能中的可解释性、泛化性及新型算法开发提供了坚实的理论支撑。未来的前沿研究将进一步探索数学与人工智能各层面之间的深度融合，推动理论与实践的共同进步。

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. Pereverzyev, S. V. An introduction to artificial intelligence based on reproducing kernel Hilbert spaces. *Springer Nature* (2022).
6. Zhou, D. X. Universality of deep convolutional neural networks. *Appl. Comput. Harmon. Anal.* **48**, 787-794 (2020).

2. 优化

优化是人工智能的核心驱动力之一，贯穿于模型训练、参数调整、性能提升等整个过程，并广泛应用于人工智能的各个分支，如监督学习中的参数优化，无监督学习的结构发现，强化学习中最大化策略的累积回报，以及自然语言处理中大语言模型的分布式优化等。

机器学习中的大多数任务可以被建模成一个优化问题，然后通过设计算法寻找好的参数以实现模型在特定任务上的良好表现。因此随着人工智能的蓬勃发展，最优化方法也得到了广泛研究。利用梯度信息的随机梯度法和小批量随机梯度法等是最基础也是最重要的一类方法，适用于大规模数据集的训练。进一步地，带有动量的随机梯度法在参数更新中通过累积历史信息，能够加快收敛并降低方差，也能在一定程度上避免局部解；典型的算法有 AdaGrad、Adam 等。更为复杂的一些自适应优化算法，则利用二阶信息或者正则化技术以实现更快、更稳定的收敛。这些方法在运行过程中，还可以巧妙结合概率统计方法，以应对训练过程中的噪声、数据分布偏差及其他不确定性因素。此外，遗传算法和粒子群优化等启发式算法在解决某些高维非凸问题时展现了较好的鲁棒性，贝叶斯优化则成功应用于深度学习中的模型超参数调优、机器学习算法的自动化调参，以及强化学习中的策略优化等场景。理论方面，对于凸优化问题，大多数梯度类算法能够收敛到全局最优；而对于机器学习中更常见的非凸优化问题，在光滑性、弱凸性或 PL 条件等假设下，能够建立收敛到稳定点甚至全局最优点的理论结果。优化动力学、神经正切核以及隐式正则化等理论则试图揭示深度学习在实践中取得成功的原因，并探究其泛化性的理论保障。对于大规模问题的分布式优化算法，不仅需要关心其计算复杂度或者迭代复杂度，通信复杂度对提高计算效率也有至关重要的影响。

人工智能的发展对优化也有促进作用，特别是大语言模型强大的生成能力。传统的优化建模依赖专家经验，而大语言模型能够利用专家知识，将自然语言描述的业务问题转化为优化模型，并调用求解器进行求解。例如，在物流路径优化任务中，模型可自动识别关键变量、约束条件和目标函数，并生成混合整数规划模型。然后通过调用 Gurobi、CPLEX 等求解器，直接生成可执行的优化代码，实现对实际场景任务的高效求解。此外，大规模求解器中往往有许多启发式规则，而如何定义这些规则并无明确标准。传统的方法一般是人工编写，或者自定义搜索空间然后进行调参，这两种方法都费时费力，并且效果局限于所定义的搜索空间。相比之下，大语言模型可以借鉴已有知识，编写许多非常高效的启发式规则，并极大提升算法性能。例如，在旅行商、车辆路由等问题中，常用遗传算法中的启发式规则可以使用大语言模型进行编写；而在可满足性问题中，主流 CDCL 算法中的启发式规则也可以使用大语言模型进行改进。

总的来讲，优化在人工智能的发展过程中起着不可或缺的作用，而人工智能的发展反过来又能促进优化建模以及优化算法的设计。不过尽管当前已经取得了非常瞩目的进展，未来仍面临许多挑战与机遇，前沿科学问题包括如下几个方面。首先，结合实际计算架构发展高效的算法是一个需要不断与时俱进的研究主题。当前优化方法以一阶算法为主，而能否发展高效的二阶算法同样值得探究。对于一些特定的问题，比如大模型中长序列、稀疏奖励下的强化学习策略优化问题，如何设计高效的方法仍然还有很大的空间。另外，尽管大模型已经初显对优化建模与算法设计的促进作用，但是如何设计运作可控的机制还需要更加深入的研究。 理论方面，机器学习中优化算法的泛化性研究在理想情形下已经有了一些进展，但是离对实际应用产生有效的促进作用仍有很长的路要走。

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).

©Just_Super / E+ / Getty

3. 统计

统计学作为数据科学的基础，为人工智能的发展提供了重要的理论支撑和方法工具。它通过数据分析、模型构建和优化算法，帮助人工智能系统应对不确定性、提取特征并实现高效决策。统计学中的收敛性质分析和统计推断，为人工智能模型的可解释性和可靠性提供了坚实理论。概率论和信息论在解释模型不确定性、构建优化理论上具有决定作用；而线性回归、广义线性模型与高维数据建模方法等统计工具，则成为机器学习和深度学习构建模型的重要基石。此外，统计学在数据预处理、特征选择与模型评估中也发挥着指导作用，为人工智能性能提升提供保障。与此同时，深度学习、生成模型、强化学习等现代技术的兴起，又使统计学与人工智能的融合催生出新的范式和问题。

深度神经网络作为人工智能的重要工具，其统计学收敛性分析一直面临挑战。现阶段，

深度神经网络仍处于经验阶段，其内在机理和理论证明常被视为“黑箱”。因此，如何基于统计理论刻画神经网络的收敛速度最小化是目前该领域的前沿问题。可借助非参数回归理论，以最小二乘或凸损失构建收敛速率；同时，针对实际中数据的非独立同分布、时序与厚尾问题，还需进一步研究收敛性能。对于扩散模型、生成对抗网络等生成模型，其统计性质的理论基础仍较薄弱。目前相关的前沿科学问题包括两个方向：一方面，如何评价生成模型估计无条件分布的效果，是一个重要课题。可基于沃瑟斯坦距离等赫尔德类概率度量手段，建立分布估计误差界，评估生成器效果；另一方面，尽管生成对抗网络在高维图像、自然语言分析等无条件分布学习上已取得进展，但在条件分布生成上仍存不足。探索条件生成器时，可从条件插值角度出发，研究插值过程的充分条件，确保条件漂移函数和分数函数在边界点处稳定，并借助沃瑟斯坦距离和 KL 散度构建误差界；或通过条件抽样，即对参考分布样本进行合适变换，利用 KL 散度匹配联合分布，再由神经网络进行非参数估计。

强化学习作为人工智能的重要分支，其收敛性和统计推断问题也极具挑战。该领域当前面临三个前沿科学问题：第一，在离线数据下如何构建统计有效的策略优化方法。一个可行技术路径为采用值增强方法，提升强化学习算法对给定初始策略的性能估计，并分析最优策略的收敛性及值差距估计量的有效性。第二，强化学习中的统计推断和假设测试问题，如在商业 A/B 检验场景下，将强化学习用于因果推断时，需考虑数据动态更新的检验方法。基于马尔可夫决策过程刻画处理与结果的时变关系，通过比较价值函数差异构建顺序检验过程，从而分析检验水平和功效。第三，无限视界环境下，策略值的置信区间的构造方法，可针对与策略相关的动作值函数建模，利用其渐近正态性构造区间。

人工智能模型“黑箱”特性一直是制约其解释性的重要因素，统计因果推断有望为此提供新的突破口。机器学习和深度学习方法在高维估计中常借助正则化降低方差，并用过拟合来部分抵消正则化带来的偏差，正则化偏差和过拟合现象也可能使估计量产生偏差，从而影响因果效应的准确推断。因此，基于统计理论的因果推断去偏方法研究是目前的前沿科学问题。可基于半参数理论构建相应框架，利用奈曼正交性与交叉拟合方法进行去偏估计，可使因果效应估计达到渐近正态分布。同时，去偏机器学习需要估计未知的里斯表示，其回归估计的有限样本均方误差及渐近性质也是值得讨论的前沿问题。

统计学在人工智能发展中扮演着不可替代的角色，而人工智能技术的进步反过来也在推动统计方法的革新。复杂模型如深度学习通过多层非线性结构能够自动捕捉高维数据中的交互效应，集成学习则为数据建模、预测的稳健性能提供保障。同时，针对海量非结构化数据，自然语言处理技术可将文本生成语义嵌入向量，转化为结构化输入以供统计建模；贝叶斯神经网络通过引入概率权重，为高维参数估计提供不确定性量化支持。随着新理论和方法的不断发展，统计学能够更高效地应对人工智能中的前沿问题，推动人工智能技术迈向更高水平，而人工智能的不断突破也将反哺统计学，使其变得更加智能与高效。人工智能与统计学之间的双向互动与协同进化，必将在未来智能革命中发挥关键作用，为社会智能化转型注入持久而强大的动力。

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. 科学计算

随着上世纪的计算机出现，科学计算迅速发展，在天气预报、油田勘探、药物设计、金融分析等领域取得巨大成功，成为理论、实验研究之外的重要支柱。但许多科学和工程领域仍存在模型不完善、机理不清晰或缺乏数学描述的问题，而传统算法在处理高维、非线性问题时常受维数灾难等限制。为解决这些难题，借助人工智能驱动的机器学习辅助建模，已成为提升复杂系统建模效率和优化计算流程的重要方向，同时推动跨学科融合与创新。

另一方面，以深度学习为代表的人工智能技术内部机理和数学理论尚不成熟，其算法稳健性和精度缺乏严格论证。大规模神经网络训练对算力需求激增，催生出“以算力替代算法优化”的趋势，形成算力与创新相互促进的“暴力”发展模式。高性能计算成为推动人工智能基础研究和应用落地的关键，人工智能也正成为新一代 E 级计算的重要引擎。

人工智能与科学计算的高度融合，前沿科学问题包括：如何设计高效数值算法加速人工智能模型；如何在科学计算工具链中创新性地引入人工智能；以及如何利用人工智能为数值模拟范式带来革新。

在高效数值算法与模型优化方面，混合精度训练和模型蒸馏压缩是当前推动模型加速部署的关键科学问题。混合精度训练利用低精度与高精度相结合，根据任务动态调整，依赖数值稳定性和误差控制方法，提高计算效率和内存利用率；而模型蒸馏通过知识传递和简化网络结构，在高维参数空间中寻找近似最优解，依托矩阵分解、张量压缩与数据量化等技术，实现压缩后模型依然具有良好预测精度，从而满足资源受限场景的需求。

在科学计算工具链创新方面，现有人工智能计算主要基于梯度下降、矩阵运算等数值方法，但在处理病态矩阵等问题时易引发不稳定性，且难以解释。近年来，学界尝试将符号计算与数值方法结合，比如采用符号搜索优化矩阵乘法、利用神经符号计算融合数据驱动与逻辑推理优势，提升模型的可解释性和泛化能力。此外，还出现了自适应多尺度计算框架，如基于图神经网络的跨尺度建模和基于 Transformer 的多尺度学习方法，这些新技术有望突破传统局限，推动智能计算向更稳定、透明、泛化的新阶段迈进。

在数值模拟方面，传统有限差分、有限元、蒙特卡洛方法在处理高维、非线性问题时常因计算负担重、收敛慢而受限。利用深度学习、强化学习等人工智能手段，研究者提出新的解决思路。深度神经网络已被用于流体力学、热传导、结构力学等领域，通过将物理定律嵌入网络，实现高效逼近。同时，人工智能促成了数据驱动与传统模拟融合方法的出现。这类融合算法先利用数据对模型降阶，再结合经典数值算法求解，既保证物理可靠性，又显著提升了计算效率，为实时监控和控制复杂系统奠定了基础。未来，随着人工智能与物理建模、数据科学和计算数学等的深度交叉融合，数值模拟将朝向更高效、智能和实时化方向发展，为复杂系统建模与优化提供精确高效的方案，推动工程实践和科学探索。

总之，人工智能与科学计算正迈入全新发展阶段，但仍面临分布式模型通信效率、跨尺度数据融合、大规模优化的稳定性等问题。这些前沿挑战的解决将促使二者进一步深度融合，可能颠覆传统深度学习框架，催生更高效、稳定、可解释的新一代计算方法与技术。同时，人工智能算法必须与底层硬件（如类脑、光计算、量子计算等新型架构）深度匹配，才能充分释放高性能计算机的算力，为大规模模型训练和推理提供保障，从而持续推动人工智能与科学计算协同进步。

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. 复杂系统

非线性科学和复杂性研究涉及多学科交叉，融合了控制论、信息科学、大数据、人工智能等智能技术，通过发展系统理论、方法和模型等，理解多变量、多层次、多尺度的相互作用，揭示其动态行为和演化规律，应用于生命、自然、社会、工程、经济、管理等复杂系统领域。前沿科学研究，正向极宏观拓展、向极微观深入、向极端条件迈进、向极综合交叉发力，正在不断突破人类认知边界。极综合交叉的核心特征是复杂性，只有在非线性与复杂性共性科学原理方面有新突破，才能从根本上解决极综合交叉问题。

人工智能、大数据、超算等新一代信息技术飞速发展，使得我们能够从多个系统和学科方向维度探索智能复杂体系的特征、智能涌现和牵引调控，为解析非线性科学和复杂性难题提供了全新的有力工具。2024 年诺贝尔物理学奖和化学奖都授予了与人工智能研究相关的科学家，以表彰他们在人工智能推动多领域基础学科研究中取得的突破性进展。人工智能的发展也依赖非线性科学和复杂性研究，来破解数据、算力需求及模型可解释性、收敛性、鲁棒性等难题，优化核心算法和架构，推动通用人工智能的变革。非线性科学和复杂性研究的发展将极大程度地反哺人工智能，推动下一代人工智能发展，即“复杂系统借力人工智能，人工智能面向复杂系统”。

围绕非线性科学与复杂性研究的三个层次，即认识、驾驭、控创，以基础数学理论为驱动力，创新发展与人工智能相互赋能的研究范式，探索不同领域复杂系统共性理论原理及共性演化调控机制，推动新一代人工智能的技术突破，应对人类可持续发展面临的重大挑战具有重要战略意义。

前沿科学问题及其突破路径包括：

第一，如何揭示各类复杂系统的共性科学原理和演化过程中所遵循的数学物理规律，并建立、整合复杂系统的基础理论框架？突破路径可以围绕形成演化、动态结构等方向展开：探索涌现行为与临界状态的普适规律、解析自组织过程与结构序的形成机制、揭示非线性与随机性驱动下的动力学演化规律、研究高维动力系统的演化特性、构建多层次复杂系统的结构演化理论、挖掘信息处理机制与智能结构的演化规律等。

第二，如何构建复杂系统要素的动态平衡性与系统结构和功能关系的普适性规律模型，并融合人工智能技术实现复杂系统精准调控？突破路径可以围绕复杂系统智能表征、动态模拟、优化调控等方向展开：研究数据驱动的智能建模与调控、融合创新多模态统计物理方法、探索多尺度动力学过程建模与调控机制、发展适用于复杂系统模拟的相场建模技术等。

第三，如何以复杂系统理论 / 模型 / 算法解决人工智能对数据和算力的巨大需求及模型的可解释性、收敛性、鲁棒性等问题，设计新的核心算法或架构以实现通用人工智能？突破路径可以围绕人工智能与复杂系统交叉融合的范式创新与算法机制等方向展开：探索基于复杂系统动力学机制的神经网络演化理论框架、构建具备结构可解释和逻辑推理能力的多层次因果模型与图网络模型、发展异质信息融合与动态知识表示方法等。

综上，人工智能等新型智能技术为解析非线性科学与复杂性难题提供了全新的有力工具；非线性科学与复杂性研究是解决新时代极综合交叉科学问题的根本途径，是设计新的核心算法或架构以发展下一代人工智能的突破口。

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).

第四章

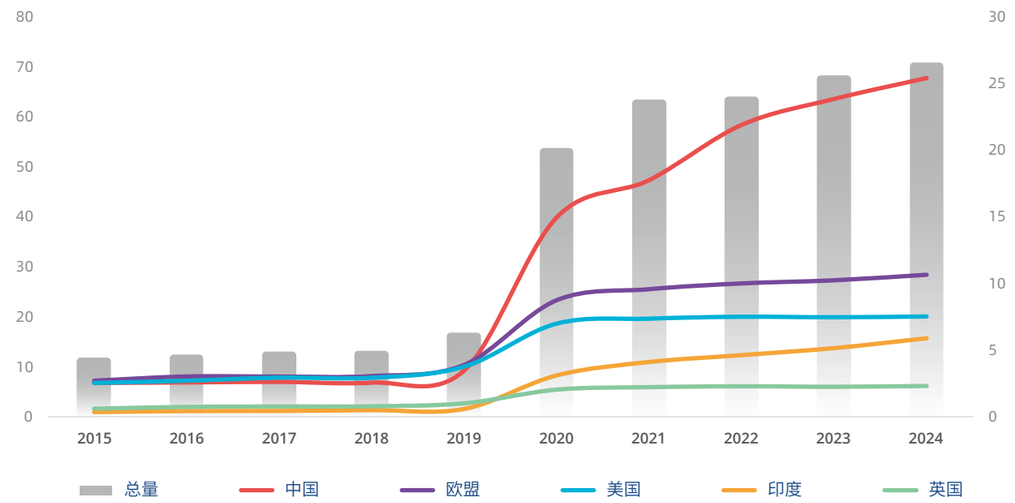
物质科学

©Xuanyu Han / Moment / Getty

AI 与物质科学

物质科学领域 AI 出版物爆发式增长始于 2020 年，出版物总量于 2024 年已达 7.07 万篇（图 4）。从国家趋势看，中国崛起和印度追赶仍是主流。数据分析和关键词词云显示，材料和电池技术主题最被关注。多尺度建模、符号回归、图神经网络和逆向设计等方法最被这一领域科学家看重，而物理启发的深度学习等方法加速了物质科学和 AI 的深度融合。

图4｜物质科学领域AI出版物总量、国家趋势（单位：千篇）与关键词词云（2015-2024）



1. 物理

1.1 背景

物理学，作为自然科学的基础学科，长期以来依赖于理论建模、实验验证和计算模拟。然而，随着研究问题的复杂性和数据规模的指数级增长，传统的研究方法在效率和精度上面临瓶颈。人工智能（AI）的兴起为物理学提供了全新的工具和思路，尤其是在数据驱动的模式识别、复杂系统建模和高效算法优化等方面展现了强大的潜力。AI 与物理学的交叉研究旨在通过将人工智能技术与物理学研究相结合，发现新物理规律、优化实验设计、加速材料发现和推动理论创新。在计算物理、量子物理、天体物理和生物物理学等领域，AI 已经显示出革命性的应用潜力。

1.2 最新进展

1.2.1 计算物理的范式革新

基于图神经网络 (GNN) 等模型的应用，已取得显著进步，代表成果包括：DeepMind 团队开发的 "GNoME" 框架，成功预测了超过 200 万种新型稳定晶体结构 [1]。DeepMind 团队提出的基于自然激发态的变分蒙特卡洛 (VMC) 方法，实现了对量子激发态波函数及可观测物理量的高精度计算 [2]。复旦大学团队开发的 AI 大模型实现对电子哈密顿量的精准预测 [3]。

1.2.2 大科学实验的智能解析和操控

基于 Transformer 架构结合对比学习等方法的突破，标志性工作包括：普林斯顿大学团队借助 AI 控制托卡马克装置中等离子体的稳定性 [4]；DeepMind 团队在托卡马克中实现等离子体的多维自主磁场操控 [5]。DeepMind 团队推出阿尔法量子比特 (AlphaQubit) 人工智能解码器，显著推动了量子纠错技术的进步 [6]。多伦多大学和加州大学伯利克分校合作团队构建 AI 模型分析射电望远镜的超大型数据集，用于寻找宇宙地外文明信号 [7]。

1.2.3 物理定律的符号化发现

基于符号回归与因果推理等模型的最新成果，代表性工作包括科学家利用 AI 技术发现了隐藏的物理对称性 , 有望拓展新物理规律的发现 [8]。麻省理工学院团队开发了物理启发的神经网络模型“AI-Feyman”，能够识别物理学公式并辅助理论推导 [9]。普林斯顿大学团队利用符号回归算法修正天文物理方程，显著提高了星系质量预测的准确性 [10]。

1.3 前沿科学问题和突破路径

当前该领域研究聚焦的三大前沿方向：

1.3.1 物理先验与神经网络的认知对齐

可微分物理定律的逆向编码，将复杂物理原理转化为可微分的约束条件；认知偏差的量化度量，评估神经网络对物理概念的 " 理解深度 "；多层级先验协调机制，解决第一性原理与唯象规律之间的冲突问题。

突破路径: 构建开放的物理学数据平台，提供高质量的训练数据；推动硬件与算法的协同创新，提升计算效率和应用能力。

1.3.2 多尺度物理系统的涌现规律挖掘

构建从微观自由度到宏观序参量的跨尺度因果涌现映射；设计具备时间尺度不变性的神经网络实现特征时空尺度的自主识别；多模态耦合的临界预测，在相变临界点附近建立基于微观涨落的早期预测指标。

突破路径：开发多尺度建模方法，实现从量子到宏观的多级关联分析。

1.3.3 自主物理发现的闭环系统构建

假设空间的几何化表示，将物理理论转化为微分流形上的路径积分优化问题；贝叶斯实验设计范式革新，构建兼具信息熵最大化和物理可解释性的主动学习策略；自动符号回归的元理论生成，建立基于拓扑数据分析的元定律发现框架。

突破路径：引入物理约束的 AI 模型，提升模型的解释性和科学性。

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).

2. 化学

2.1 背景

随着数据科学和机器学习技术的发展，AI 为化学研究带来了新的机遇，不仅深度探索了化学相关的专业问题，还辐射到了多个相关领域。当前，AI 与化学的交叉研究主要应用于物质科学、模拟计算、实验流程创新和化学生物学等领域 [1]。

AI 技术从理论和实验等方面出发推动了化学领域发展。理论方面，基于 AI 的新算法显著减少了所需的计算资源，提升了模拟的效率。实验方面，AI 推动了实验的高度自动化，提供了更加实用和高效的解决方案。未来的研究将继续深入发掘 AI 在化学中的潜力，推动科学技术的发展。

2.2 最新进展

2.2.1 AI+ 机器人化学家

AI 机器人化学家结合了移动机器人、分析仪器等硬件和自动化控制软件，在化学合成中实现更广泛、更灵活的自动化实验与决策过程。多项研究加入了由 AI 提供的化学见解，在人机协作下开展实验，提高了数据处理的效率和准确性，加速研发进程 [2,3]。

2.2.2 AI+ 势能面描述

基于原子模型的人工神经网络能构建包含能量、力、应力等信息的势能面，在确保计算精度的前提下实现模拟过程的加速 [4,5]。全局神经网络势函数方法被应用于反应过渡态搜索、反应机理预测等方面，提高复杂体系预测能力 [6,7]。

2.2.3 AI+ 结构设计与大数据平台

基于大语言模型的化学语言模型在相关领域也有所应用，例如将材料结构或药物分子转化为可识别的信息后进行性质预测或结构设计 [8,9]。此外，图神经网络、生成模型等算法也能从现有结构出发进行新分子、新材料的设计 [10]。这些进展促进了融合大数据与人工智能的研究平台的诞生，用户可以通过对话等形式获取知识库中收录的信息，辅助实现快速高通量筛选 [11]。

2.3 前沿科学问题和突破路径

2.3.1 针对化学 AI 模型的效率优化

化学现象可能涉及从电子、原子到宏观的不同尺度。物质的丰富性和反应的多样性决定了 AI 模型的训练和数据集的扩充需消耗大量计算和存储资源。

突破路径：开发高效的算法和模型用于不同尺度模拟过程。针对化学数据的特点，设计高效的数据库管理系统，用于高效存储和快速查询。

2.3.2 通用化学数据集的建立

化学物质结构多样，实验条件参数繁杂，数据来源可能拥有计算或实验等多重来源。需要注重数据质量，确保选取数据时的公平性和完整性。

突破路径：统一化学数据的标准格式，例如分子式使用 SMILES 表示，物理量单位统一使用国际单位制。开发自动化工具对来自化学实验、理论计算结果、文献资料等多重来源的数据进行清洗、验证和标注。

2.3.3 增强 AI 模型的化学特化型认知

此前大语言模型大多是基于语言构建的，而针对化学领域的 AI 模型需要在解决更复杂问题时具备更强的化学科学认知，例如需要正确理解分子和晶体结构、具备基本的化学常识。

突破路径：结合知识图谱等方法，加入化学领域的基本科学定理和物质数据信息，解决大语言模型现存的短板 [12]。强调“正在解决化学领域科学问题”的系统提示词，以确保调用上述知识背景。

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. 材料

3.1 背景

近年来，材料科学的发展面临诸多瓶颈，严重制约新材料的高效开发与产业化。首先，传统研究依赖经验驱动的实验与理论计算，从概念验证到应用往往需十年以上时间。与此同时，材料研究高度依赖昂贵的实验设备和计算模拟，尤其在锂离子电池、电催化、先进高分子等领域。此外，材料的结构与性能涉及电子、原子、介观至宏观多个尺度，传统方法难以构建高效的跨尺度建模框架并处理复杂的材料空间。面对这些挑战，人工智能与材料科学的深度融合，有望大幅缩短研发周期、降低成本，并加速材料科学向智能化迈进。

3.2 最新进展

3.2.1 新材料的设计

当代新材料设计正从传统高通量筛选向基于机器学习的逆向设计加速发展。生成式模型突破了材料搜索空间的限制，成为新材料发现的核心技术。主要方法包括变分自编码器、生成对抗网络、扩散模型和大语言模型（LLM）等。例如，MatterGen[1] 利用扩散模型生成稳定材料。

3.2.2 材料性质预测

当前材料性质预测广泛采用图神经网络、多保真度学习与深度学习 Hamiltonian 等方法，以兼顾高精度与大规模搜索需求。在电池、催化、高分子与光电材料等前沿领域，这些方法正被深度耦合，以快速筛选高价值材料。

3.2.3 材料科学 AI 智能体和大模型

材料科学 AI 智能体和大模型通过图神经网络、LLM 等技术，实现关键性能的精准预测及科学推理的可靠性。例如，MatChat AI 智能体 [2]，依托论文知识库，实现可追溯的材料科学问答。MatterChat 材料大模型 [3] 则通过桥接模型将高分辨率的原子结构与 LLM 文本表示融合，在提升预测精度的同时，也提供了更友好的人机交互方式。

3.2.4 自主实验室实现材料智能合成

材料智能合成正借助大规模密度泛函理论（DFT）计算数据库、生成式结构预测与 LLM 快速发展。机器人自动化与云端调度协同优化实验流程，确保模型实时更新。例如，自主实验室 A-Lab[4]，利用 DFT 数据库与自然语言模型驱动无机粉末合成闭环；云端异步分布式协作模式 [5]，可以实现实验平台在中央 AI 调度下独立运行，大幅提升研发效率。

3.3 前沿科学问题和突破路径

3.3.1. 构建可靠的、多源异构材料数据库，并推动全球数据共享。

材料科学领域数据资源的多源异构性、数据质量参差不齐、数据共享受限等问题导致 AI 模型泛化能力受限，降低了材料预测的可信度。

突破路径：构建高可信度的多源异构材料数据库需要统一数据存储格式，确保计算、实验和 AI 模拟数据的跨平台兼容；其次，构建多保真度数据集，通过清洗、补全和去重提高数据质量。在数据共享机制上，构建开放材料大数据生态，推动高质量数据共享。

3.3.2. 开发具有化学可行性且可扩展的生成式 AI 材料模型，以实现多尺度自主材料设计。

现有生成式 AI 在材料科学中虽然能够生成新材料，但忽略了热力学稳定性、合成路径和器件级别的性能关联，难以真正实现材料从原子级发现到宏观系统优化的全面整合。

突破路径：构建兼具化学可行性与高灵活度的生成式 AI 材料模型，需要结合物理约束生成模型，并引入 DFT 等筛选机制提升热力学稳定性。同时，结合 LLMs+ 自主实验室优化合成路径，实现实验可行性评估。

3.3.3 将 AI 与自主实验室集成，实现材料发现全自动化及性能智能优化。

高效整合 AI 计算、实验自动化与多尺度建模，加速设计室温超导材料、自修复柔性半导体材料、超轻高强纳米复合材料等革命性材料已成为推动材料科学变革的前沿问题。

突破路径：自主实验室的突破依赖多层次 AI 集成：首先，利用生成模型生成符合目标性能的候选材料并进行精准筛选；其次，依托机器人自主合成与智能表征，实现材料制备全自动化；最后，通过强化学习与自适应优化，持续优化材料设计。

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. 能源

4.1 背景

随着全球能源需求不断增长以及环保压力的加剧，传统科研模式在提升效率和推动能源材料创新面临巨大挑战。现有的能源材料研究进展缓慢，且已接近理论极限。以催化类材料为例，即使机器人可以协助实验合成，并把一次合成、表征和测试压缩至一周的时间，仍需 100 多年才能筛选出 5,000 种潜在组合 [1]。考虑到验证周期和耗费成本，难以满足对低碳、廉价、高效新型能源材料的需求，特别是在实现“双碳”战略目标下，迫切需要寻找一系列突破性的新型能源材料设计。为应对这一需求，AI 驱动的能源材料研究为解决这一瓶颈提供了新思路。通过人工智能的辅助，可以加速分子材料的设计、发现与优化，大幅提升能源材料的研发效率 [2]，助力实现 2030 年碳达峰和碳中和目标。

4.2 最新进展

在能源材料研究中，机器学习正推动储能、催化等领域的创新发展，正在从加速海量数据筛选，挖掘复杂构效关系和智能优化材料参数等多个方面推动材料研发从试错模式向 " 设计 - 验证 " 闭环范式转型。Cao 团队 [3] 开发生成式强化学习框架，通过优化分子稳定性和氧化还原电位，筛选出氧化还原液流电池的新型有机候选分子；Chen 等 [4] 采用无监督学习与化学信息学结合，从电化学特征中提取关键分子片段，构建高质量电池数据库并成功设计出锂离子载体分子；Wu 团队 [5] 将无监督学习与爬坡微扰弹性带模拟结合，实现固态电池中锂离子导体的快速筛选，基于小规模计算数据库建立了高导电率电解质的预测模型。Wu 等 [6] 利用贝叶斯优化匹配有机半导体高通量合成数据，开发出性能优异的太阳能电池空穴传输材料；Wu 等 [7] 提出机器学习辅助的二维钙钛矿合成框架，融合实验数据与化学先验知识，显著提升新材料合成效率；Angello[1] 创新性提出闭环转移研究方法，通过贝叶斯优化迭代优化分子光稳定性，结合机器学习验证假设指导材料发现。Hu 等 [8] 建立机器学习驱动的高通量筛选协议，结合光敏化、电子转移与催化描述符，优化出高效分子光催化 CO_2 还原系统；Zhong 等 [9] 构建 ML-DFT 协同反馈框架，通过机器学习表面电负性、CO_2 吸附能等关键指标，设计出低能垒催化结构；Li 团队 [10] 整合二维 - 三维机器学习算法，建立覆盖全 C-C 偶联前体与活性位点的数据集，实现复杂电催化偶联反应机理的快速解析。

4.3 前沿科学问题和突破路径

4.3.1 数据壁垒制约研发效率

能源材料多源异构数据（实验、物性、行为）分散且标准缺失，跨领域数据整合困难，阻碍大规模数据库构建与知识迁移。应当建立“微观结构 - 宏观性能”标准化评价系统，通过物理模型与数据驱动融合，提取可迁移的结构特征描述符。搭建覆盖电化学特性、催化活性等核心参数的多维度验证平台，形成“计算预测 - 实验验证 - 模型迭代”闭环研发链条。

4.3.2 模型黑箱限制机理认知

深度学习预测结果与微观机制脱节，难以解析构效关系，制约新材料设计的理论指导性。在模型内部嵌入基本的物理化学函数约束层，开发数据 - 机理双驱动的分子生成模型。

4.3.3 多维优化陷局部最优困境

现有 AI 算法在导电性、稳定性、成本等多目标协同优化中易偏离全局最优解，限制复杂场景应用。集成能量转化—存储—利用场景的关键宏观指标与微观特征，构建跨介质、跨尺度的能源分子数据库。开发具备物理约束的生成模型，在化学空间内实现多目标帕累托前沿搜索与全局优化。

4.3.4 实验和设计闭环尚未有效贯通

AI 预测与实验验证间反馈迟滞，高通量计算与自动化实验平台协同不足，迭代效率亟待提升。突破传统线性研发路径，建立“假设生成 - 智能筛选 - 定向合成”的跨维度研发体系，通过迁移学习实现小样本场景下的知识复用。

4.3.5 安全评估体系缺失

环境毒性、长周期稳定性等安全指标缺乏量化标准，AI 优化框架未充分融合可持续性约束条件。将安全性评估前置化嵌入 AI 设计流程，融合生命周期分析与材料基因工程，开发兼顾性能与可持续性的多目标优化算法。

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).

第五章

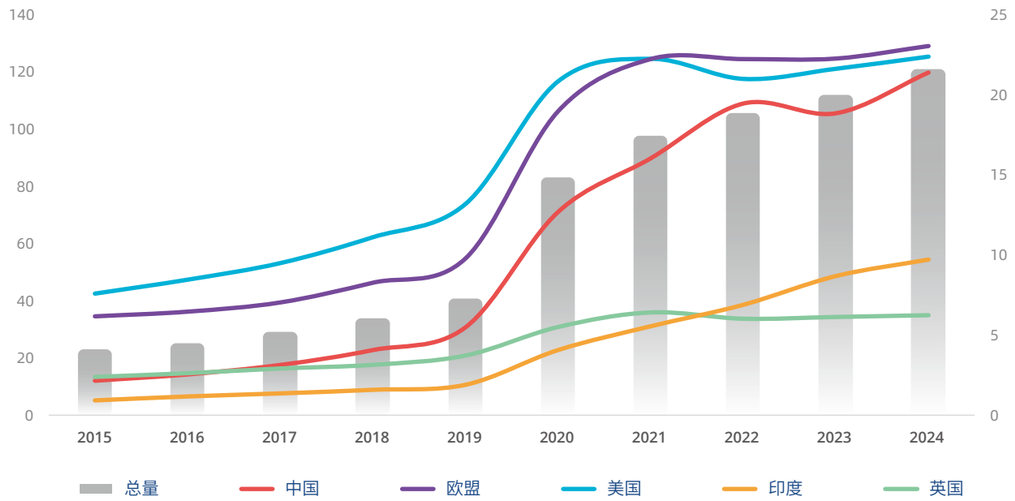
生命科学

©Andrew Brookes / Connect Images / Getty

AI 与生命科学

生命科学领域 AI 出版物的爆发式增长始于 2020 年，2024 年全球出版物达到 12.07 万篇（图 5）。在美国与欧盟长期主导该领域研究的背景下，中国加速追赶，2024 年出版物总量已接近欧美水平。数据分析和关键词词云显示，脑神经、基因和健康领域的研究最被关注；AI 通过大语言模型加速药物研发，通过高分辨率成像重构疾病诊断，多尺度数据驱动模型和基因组学、蛋白质组学深度融合，解码生命复杂系统的内在规律。

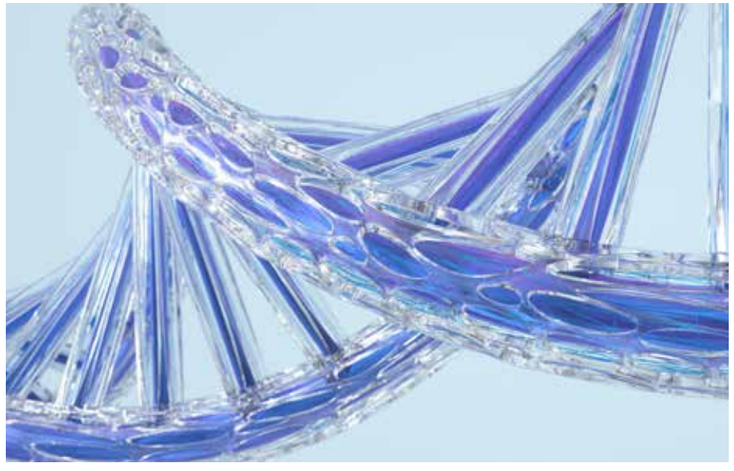
图5｜生命科学领域AI出版物总量、国家趋势（单位：千篇）与关键词词云（2015-2024）



1. 合成生物学

1.1 背景

合成生物学是一门深度融合了生物学、工程学、计算科学等基础和应用学科的新兴交叉学科，旨在设计、改造乃至重新合成新的生物系统，实现特定功能。随着基因编辑、蛋白质从头设计、代谢工程等技术的进步，合成生物学近年来蓬勃发展，已在医药健康、环境治理和新型材料等众多领域产生变革性影响。人工智能（AI）技术，凭借其强大的多任务学习能力和未知空间智能探索能力，有效地契合了合成生物学的智能化设计需求，并为破解生物系统序列—结构—功能之间的复杂映射关系开辟了全新路径，推动合成生物学向更高效、更精准的方向快速发展，引领其从“生物改造”迈向“生物创造”时代。



©Andriy Onufriyenko / Moment / Getty

1.2 最新进展

1.2.1 人工智能赋能基因编辑和核酸疫苗

以 CRISPR-Cas9 系统为代表的基因编辑和以 mRNA 为核心的核酸疫苗，在 AI 驱动下，正在经历一场深刻的技术革新。通过深度学习与大规模数据分析，AI 能够从复杂的基因组数据中识别治疗靶点，并精确预测基因编辑和抗原的生物学效应，显著提升分子调控的精确性与效率[1]，实现对细胞功能的精准编程，为个性化治疗和精准医学提供创新的解决方案。

1.2.2 人工智能驱动蛋白质从头合成

蛋白质作为生命功能的核心执行者，其设计与合成在合成生物学中占据重要地位。AlphaFold 实现了蛋白质结构预测的革命性突破，其预测精度达原子级别，覆盖 98.5% 人类蛋白质组[2]。基于扩散模型的 RFdiffusion 蛋白质生成工具[3]，为蛋白质从头设计提供了全新路径。基于自然语言模型的 ProGen[4]、EVOLVEpro[5] 等工具，实现了酶、治疗性抗体等特定功能蛋白从头合成和定向进化，展示了 AI 在全新功能蛋白设计中的强大潜力。

1.2.3 人工智能重塑生物智造

人工智能与微生物基因组学的深度融合，为合成生物学提供革命性工具。通过 AI 驱动的数据分析和模型优化，精准确定并优化特定化合物的合成路径，针对性地改造菌株基因组，实现青蒿素、萜类化合物及其衍生物的高效生产。生物 AI 模型 Evo，能够生成具有特定功能的全新 DNA 序列，甚至设计完整的微生物基因组[6]，促使合成生物学从局部优化迈向全局设计的重大跨越，为建立高效定制化微生物工厂开辟新路径。

1.3 前沿科学问题和突破路径

1.3.1 复杂生物系统的设计与可扩展性

单个生物分子（如基因、蛋白质、脂质）的从头设计已取得进展，但复杂生物系统（如纳米机器人、真核细胞）的设计与拼接仍面临挑战，且缺乏可扩展性，难以满足大规模生产需求。

突破路径：

开发 AI 驱动的模块化设计工具，实现复杂生物系统自动化拼接与优化。开发基于进化算法的适应性优化模型，提升系统的可迭代性与可扩展性。

1.3.2 治疗性功能蛋白的定制化设计与应用

尽管 AI 能够设计全新功能蛋白，但其在体内易被免疫系统识别攻击，且受限于蛋白 - 蛋白互作的 AI 预测精度不足，在疾病治疗中尚未突破。

突破路径：

创新 AI 算法，开发融合多模态数据的深度学习模型，实现高精度蛋白互作界面预测与功能设计。合成“免疫隐形”智能蛋白，通过环境响应动态屏蔽免疫表位，实现治疗性蛋白的体内长效稳定。

1.3.3 生物分子集成电路与生物计算机器

生物计算硬件集成度低、自动化运行能力不足，限制了其在高并行计算、复杂系统优化等领域的应用潜力。

突破路径：开发基于 AI 的生物元件模块化设计工具及集成系统，提升集成电路的规模与功能。推动电子 - 分子双向通讯技术，提升计算带宽与效率，构建全自动生物计算机器。

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. 医学

2.1 背景

AI 在医学领域经历了从规则系统到深度学习的演进，应用从信息化延伸至个性化诊疗[1]。早期 AI 应用于计算机辅助诊断和临床决策支持系统，通过影像识别与数据分析提升诊疗精度。2017 年 Transformer 架构突破催生 BERT、GPT 等预训练模型，显著提升医学文本解析能力，奠定大语言模型（LLM）及多模态大语言模型（MLLM）基础。

当前 LLM 已能处理复杂医学文本，包括病历解读、文献分析及辅助决策推理[2]；MLLM 融合文本、影像、基因组等多源数据，构建跨模态诊疗框架提升诊断效能[3]。这些技术突破使 AI 医疗在资源匮乏地区显现特殊价值，通过自动化分析降低医疗成本，缓解人力短缺。全球开源社区的协同创新加速医学大模型多语言适配与全球化应用，释放临床转化潜力[4]。

2.2 最新进展

2.2.1 医学 LLMs

LLMs 在医学文本解析中展现出卓越能力，可自动化解析电子病历、PubMed 文献及医学指南，精准提取药物名称、副作用等关键信息，有效支撑临床试验匹配与知识图谱构建。在交互应用方面，LLMs 驱动的聊天机器人（如 Woebot）通过认知行为疗法干预心理健康问题[5]。2023-2024 年，科技巨头相继发布医学专用 LLMs：Google 的 Med-PaLM 2 通过美国医学执照考试认证，达到专家水平[6]；Microsoft 的 BioGPT 聚焦生物医学文本生成，其关系抽取任务的 F1 分数表现突出[7]。值得注意的是，LLMs 在整合遗传变异数据提升基因表达预测精度及泛化能力方面展现独特优势，为解析复杂疾病遗传机制、筛选药物靶点提供新路径，加速药物研发。

2.2.2 医学 MLLMs

MLLMs 通过融合影像、基因与临床数据，促进疾病机制的多维解析，赋能精准医疗。北京大学开发的 MuMo 模型基于 429 例 HER2 阳性胃癌患者的多模态数据（影像学、病理学及临床信息），实现了抗 HER2 治疗与免疫治疗响应预测，凸显多模态数据在解析肿瘤异质性中的价值[8]。MedFound-DX-PA 在零样本学习下取得了 80.7% 的疾病诊断准确率，印证多模态 AI 医学建模的专业化突破[9]。目前 MLLMs 已逐步渗透临床实践，覆盖疾病筛查、诊断到个性化管理的全流程。

2.3 前沿科学问题和突破路径

2.3.1 医学大模型知识体系缺乏广度与深度的融合

需融合知识广度与深度应对医学复杂性：广度需整合电子病历、影像、基因组等跨领域数据，深度需疾病机制与个体化诊疗的专家级理解。数据隐私限制全球共享，地区标准差异与知识更新滞后制约适用性。应建立动态知识体系保障全面精准。

突破路径：利用自然语言处理 + 知识图谱构建动态体系[10]，多模态融合提升诊断精度。采用分布式学习实现隐私保护下的全球共享，建立医学知识中枢。

2.3.2 临床决策大模型缺乏证据链和推理

现有模型难提供完整证据链，结论缺乏透明推理，影响高风险场景医患信任。多模态数据融合（基因组 + 临床 + 影像 + 生活方式）需更强推理能力。应构建跨学科推理机制应对复杂诊疗。

突破路径：强化因果推理提取关键证据[11]，构建多智能体系统模拟多学科会诊。结合可靠数据源生成可视化证据链，满足监管与临床需求。

2.3.3 模型缺乏可解释性与可靠性

黑箱特性降低临床信任，限制高风险决策应用[12]。存在幻觉风险（如生成错误信息）需优化。监管要求“透明决策依据”，但可解释性与精确性存在矛盾。需建立可验证推理过程，通过可视化提升透明度。

突破路径：融合知识图谱与可视化技术（如热图定位）解释决策逻辑，开发交互式界面支持参数调整。借鉴 DR.KNOWS 模型提升诊断可追溯性[13]。

2.3.4 伦理与隐私风险

训练涉及患者影像 / 基因等敏感数据，存在隐私泄露风险。需明确 AI 误诊责任，解决资源匮乏地区技术公平性，避免加剧医疗不平等[4]。临床部署需平衡知情同意、医生培训与系统公平性。

突破路径：采用联邦学习[4]+ 差分隐私 + 同态加密构建保护体系。制定全球 AI 医学伦理准则，优先选择符合 HIPAA 法案的本地化部署[14]，通过伦理培训增强信任。

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. 神经科学

3.1 背景

人工智能（AI）与神经科学的深度融合正以前所未有的速度推动我们对大脑运作机制和智能本质的认识。神经科学通过解析大脑的结构、功能和认知模式，为 AI 提供生物启发，促进了感知网络、脉冲神经网络等算法的发展 [1]。同时，AI 利用大数据和深度学习技术加速了脑影像、连接组学及神经计算模型的构建，从而提升了神经疾病的诊断与治疗效率 [2]。近年来，通过自动化图像分割、三维重构及多模态数据融合，研究者已成功重建了从线虫到人类大脑皮层样本的复杂神经网络，展示了 AI 在处理海量脑数据方面的巨大潜力 [3,4]。这种双向赋能不仅推动了基础科学的突破，也为临床应用和类脑智能的发展奠定了坚实基础 [5]。

3.2 最新进展

近年来，AI 技术在神经科学各领域取得了显著进展，主要体现在数据采集、处理与分析上。在脑连接组学方面，高分辨率成像结合自动化图像处理技术，使得研究者能够重构从小型生物到人类大脑皮层样本的神经网络。例如，哈佛大学与 Google 合作利用连续切片电子显微镜及自动图像分割技术，对 1 立方毫米大脑皮层样本进行了三维重构，记录了约 1.5 亿个突触数据 [2,3]。此外，东南大学与美国脑计划合作，通过多模态全脑光学显微影像绘制了从亚微米到全脑尺度的神经元全景图谱 [6]。

在临床应用领域，AI 已广泛用于 MRI、PET、EEG 等多模态影像的自动分割与异常检测，显著提升了阿尔茨海默症等神经退行性疾病的早期诊断准确率 [1,7]。在神经信号解码和脑机接口技术方面，利用深度学习算法实现了瘫痪患者的想象写字、手指精细控制等运动意图的实时高效解码 [8,9]。

此外，多模态数据融合与跨尺度建模为整合解剖、功能及分子信息提供了全新视角。近期，Jiang et al. 提出的 NeuroXiv 平台实现了对全脑神经形态数据的动态挖掘 [10]；而 Liu et al. 则通过全脑形态测量揭示了神经元多尺度模式之间的内在联系，为构建生物神经网络模型提供了理论依据 [6]。同时，ARNI Institute 开发的连接组驱动神经架构搜索方法为类脑计算开辟了新思路 [11]。这些进展不仅丰富了神经科学的实验手段，也为精准医疗和智能计算的未来奠定了坚实基础。

3.3 前沿科学问题和突破路径

3.3.1 智能数据生成与标注

电子显微镜、光学显微镜和扩散 MRI 等成像技术实验生成了海量、多模态数据，如何利用生成对抗网络等 AI 技术实现数据自动生成与智能标注，从而准确捕捉神经网络微观结构和功能特征，成为当前亟待解决的关键问题。

突破路径：

构建 AI 驱动的神经信息学平台，整合实时多模态数据流，无缝融合多模态数据。开发基于生成对抗网络的跨态数据生成框架，实现不同成像技术数据的相互转换与补充。构建融合对比学习和主动学习的半自动标签系统，减少人工标注工作量。

3.3.2 新型脑机接口与高精度神经信号解码

脑机接口技术正向非侵入性与微创性相结合的方向迈进，如何通过多模态信号融合，实现复杂神经信号的实时高精度解码，推动人机交互技术的革新？

突破路径：整合多模态数据，开发自适应解码算法，实现动态个性化的稳定神经信号解码，以适应神经信号的非平稳特性。构建高阶认知状态解码框架，实现从低级运动意图到高级抽象思维（如记忆等）的跨层次语义解码，从而推动脑对脑通信或 AI 辅助的记忆增强技术的发展。

3.3.3 构建可解释、具生物启发性的 AI 模型

突破现有 AI 模型的“黑箱”局限，构建出既能高效预测又具备生物启发性和可解释性的模型，有助于揭示大脑认知过程中关键的决策机制，并为临床诊断和治疗提供更直观的理论依据。

突破路径：

开发符号 AI、图神经网络与传统深度模型相结合的混合架构，自动将深度学习预测结果映射到可理解的神经生物学概念，解释预测背后的生物学机制。构建能够预测个体对神经干预响应的 AI 驱动模型，使 AI 模型能够在无需大量重新训练的情况下，从大型数据集中高效适应个体案例。

3.3.4 多尺度多模态脑结构和功能网络模型与跨层次数据整合

信息处理贯穿从单个神经元至全脑网络的各个层面，构建涵盖细胞级、区域级和全脑级的多尺度综合模型，为深入理解高级认知功能和行为提供全新解析框架。

突破路径：

构建多层级多尺度融合的全脑脑图谱，将不同结构和功能模态的数据同步到统一时间轴上，打破当前数据融合的瓶颈。构建 " 细胞 - 环路 - 系统 " 的分层动态脑网络模型，实现不同尺度的脑信息双向传递。

上述前沿问题的研究将为新一代类脑智能系统及精准医疗技术的发展提供理论基础和实践指导。AI 正推动神经科学从实验室探索迈向临床应用，为揭示大脑复杂机制和开发新型治疗方案提供了全新视角。通过加速脑数据的获取、整合与智能解析，AI 不仅助力理解神经系统的基本原理，也为神经退行性疾病、脑机接口及类脑计算提供了坚实支持，全面解码大脑秘密和构建高效智能系统将有无限可能。

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* https://doi.org/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. 医疗

4.1 背景

AI 技术的快速发展为医疗健康领域带来了新的机遇。从早期疾病筛查到精准治疗方案设计，从药物研发到健康管理，AI 的应用正逐步覆盖医疗健康的全链条。尽管 AI 在医疗健康领域的应用前景广阔，其发展仍面临诸多技术和伦理挑战。如何平衡技术创新与隐私保护、如何确保 AI 的可靠性和透明性、如何解决跨学科协作中的瓶颈问题，都是亟待解决的关键议题。

4.2 最新进展

4.2.1 疾病诊断与预测

AI 在疾病的早期筛查、诊断与预后分析中展现出巨大潜力。深度学习在肺癌、乳腺癌等影像诊断中已实现超越放射科医生水平的准确率 [1]。Med-PaLM2 [2] 顺利通过了美国医师执照考试（USMLE），可以根据 X 光片等多模态信息对患者的病情进行分析和诊断。通用自监督病理学基础模型 UNI 利用超 1 亿张组织病理学图像训练，具有 108 种癌症亚型泛化分类能力，大大降低了对数据标注的依赖 [3]。

4.2.2 药物研发

AI 可以快速筛选出潜在的药物候选分子，并优化临床试验过程，极大缩短药物研发周期并优化成本。全球首款由生成式 AI 完成靶点发现和分子设计的候选药物 TNIK 抑制剂 INS018_055[4]，现已进入 2 期临床试验，并在 IIa 期临床试验中取得积极结果。

4.2.3 个性化医疗与健康管理

AI 可以通过整合基因组等多组学数据、生活方式数据、个体治疗数据，为个性化健康管理提供支持 [5]。智能可穿戴设备与 AI 健康助手结合，通过分析心率、血糖等生物参数，在慢病管理与健康干预中发挥重要作用。

4.2.4 公共健康与疫情监测

公共卫生领域中，AI 推动了疫情传播模型预测、疫苗分发优化、卫生政策精准实施 [6,7]。在新冠疫情期间，AI 帮助构建了实时疫情监测、传播路径预测与风险人群识别平台，支持了科学防疫政策 [8]。

4.3 前沿科学问题和突破路径

4.3.1 如何实现多模态数据与专业知识的融合

当前 AI 应用多限于单一数据源或场景，如何实现多模态数据和行业专家专业知识的有效融合，构建跨学科、多层次的生态系统仍是重要问题。

突破路径：

构建多模态联合表示框架，结合知识图谱的结构化知识增强融合表示的可解释性。开发交互式系统，支持医学专家对结果进行标注和修正。

4.3.2 如何确保模型的公平性与可信性

由于健康数据的异质性以及样本偏差，AI 在健康预测和诊断中可能引发算法歧视问题。此外，AI 模型进行健康决策过程中的不可解释性问题削弱了医学专家和公众的信任感。因此，如何构建公平性和可信性兼备的 AI 技术仍是关键挑战。

突破路径：

开发可解释的神经网络架构，帮助医学专家理解和信任 AI 决策的过程。强化算法公平性约束，建立多样化的数据标准及跨群体评估指标，确保不同性别、种族和地区患者均能获得公平的 AI 诊疗支持。

4.3.3 数据孤岛与隐私保护问题

医疗健康数据分散在不同的机构和系统中，形成数据孤岛，难以有效共享和整合。健康数据高度敏感，涉及个人隐私及伦理问题。如何在保证数据隐私的前提下实现数据共享和整合，是一个亟待解决的难题。

突破路径：

发展联邦学习技术，通过在不交换数据的情况下开发分布式模型训练技术。强化隐私保护技术，为跨国界健康数据协作提供技术支持。建设医疗 AI 测试基准与标准化认证体系，推动跨学科数据与模型共享。

4.3.4 伦理与监管挑战

AI 在医疗健康领域的应用涉及伦理问题，责任主体难以界定，现行法律体系缺乏针对性条款。AI 技术发展迅速，政府监管政策往往滞后，无法适应新的技术和应用。

突破路径：

建立针对医疗 AI 的伦理和法律框架，明确责任主体，规范数据使用和模型应用，保障患者权益。建立“沙盒监管”机制，允许创新技术在限定范围内试运行，同步收集临床反馈以动态更新法规。

AI 正在以革命性的方式推动全球健康领域的变革。AI 赋能健康领域的潜力不可估量，但它仍处于发展的初级阶段，需要技术、伦理、法规和应用的协同推进。持续推进 AI 技术在医疗领域的应用与研究，将不仅是对技术发展的推动，也是对人类福祉的深刻贡献。在未来的十年内，AI for health 或将成为提升全球医疗公平和人类健康水平的核心动力之一。

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. 演化

5.1 背景

演化 (Evolution) 或通常称为进化是理解地球生命多样性的基础性概念。数十年来，科学家依赖化石分析、基因测序和野外观察等传统方法研究进化。然而人工智能的出现彻底改变了进化研究范式。AI 处理分析大规模数据、识别复杂模式并作出预测的能力，为理解物种随时间进化的机制开辟了新维度。

与此同时，AI 自身正以指数级速度进化。AI 驱动的技术进步重塑着人类生活的每个方面，从医疗健康、交通运输到通讯娱乐。AI 与进化呈现双向互动: AI 助力生物进化研究，而 AI 驱动的技术进化也在反方向影响人类与非人类物种的进化轨迹。

5.2 最新进展

AI 在生命演化研究中的应用

1. 基因组分析

突变检测：CNN 等算法精准识别 DNA 序列中的功能突变，预测其对选择压力的响应（如适应性或中性变异）。

系统发育重建：贝叶斯推断与邻接法整合多基因数据，提升进化树准确性，解析复杂遗传关系（如物种分化事件）。

2. 表型分析

形态特征：计算机视觉量化叶片脉络、化石三维结构，揭示形态 - 环境协同进化规律。

行为特征：可穿戴设备结合 ML 算法分析动物社交行为（如灵长类理毛），关联行为模式与生存优势。

3. 进化模拟

遗传算法：模拟抗生素压力下细菌耐药性进化，优化选择压力参数。

基于主体的模型（ABM）：构建捕食者 - 猎物动态系统，解析性状演化与环境适应机制。

模型优化：EVO1 实现 7B 参数基因组基础模型，单核苷酸分辨率处理 131kb DNA 序列，基于 270 万原核 / 噬菌体基因组训练，整合中心法则模态（DNA/RNA/蛋白）与多尺度进化学习 [1]。EVO2 扩展为 7B/40B 参数模型，跨生命域训练 9.3 万亿碱基，具备单核苷酸分辨率的百万 token 上下文窗口，生成线粒体 / 原核 / 真核序列的自然度超越现有方法 [2]。

5.3 前沿科学问题和突破路径

AI 对技术进化的影响

1. 医疗健康与人类进化

选择压力变迁：AI 诊断提升疾病早期干预，削弱遗传缺陷的自然选择强度（如癌症易感性）。

能力增强：外骨骼与认知增强设备突破生物限制，可能重塑人类性状竞争优势（如学习效率）。

2. 环境与生态干预

精准农业：AI 驱动的农药施用加速害虫抗药性进化，改变土壤微生物群落结构。

机器人交互: 农业机器人影响作物生态，监测无人机干扰野生动物行为模式（如迁徙路径）。

案例

AlphaFold：通过蛋白质折叠解析揭示进化保守性，加速药物靶点发现。

进化机器人：MIT 实验室利用遗传算法开发环境自适应机器人，模拟自然选择机制。

伦理挑战

1. 研究偏见

数据偏差：基因组数据集中于特定族群（如欧洲人群），导致进化推论偏差。

算法偏差：系统发育模型假设错误可能误导生物多样性保护策略。

2. 进化干预风险

基因编辑：AI 辅助 CRISPR 技术可能引发非目标突变或生态链破坏（如基因驱动生物扩散）。

技术鸿沟：认知增强设备获取不平等或加剧社会分层，影响人类性状进化方向。

人工智能与进化科学的融合正在开创具有变革性的研究领域。AI 不仅深化了生物进化机制的理解，更重塑着跨物种的生物进化轨迹与技术发展路径。例如，科学层面：深化对适应性进化（如高海拔适应）与基因 - 环境互作的理解；技术层面：推动医疗、农业与生态保护的范式革新。当前突破性进展需要严谨伦理审视: AI 分析中的系统性偏见、进化路径干预的伦理困境，以及技术采纳不均加剧的社会经济分化。通过前瞻性治理框架应对这些挑战，方能确保 AI 的进化应用产生公平社会效益，控制潜在的生态与人类风险。

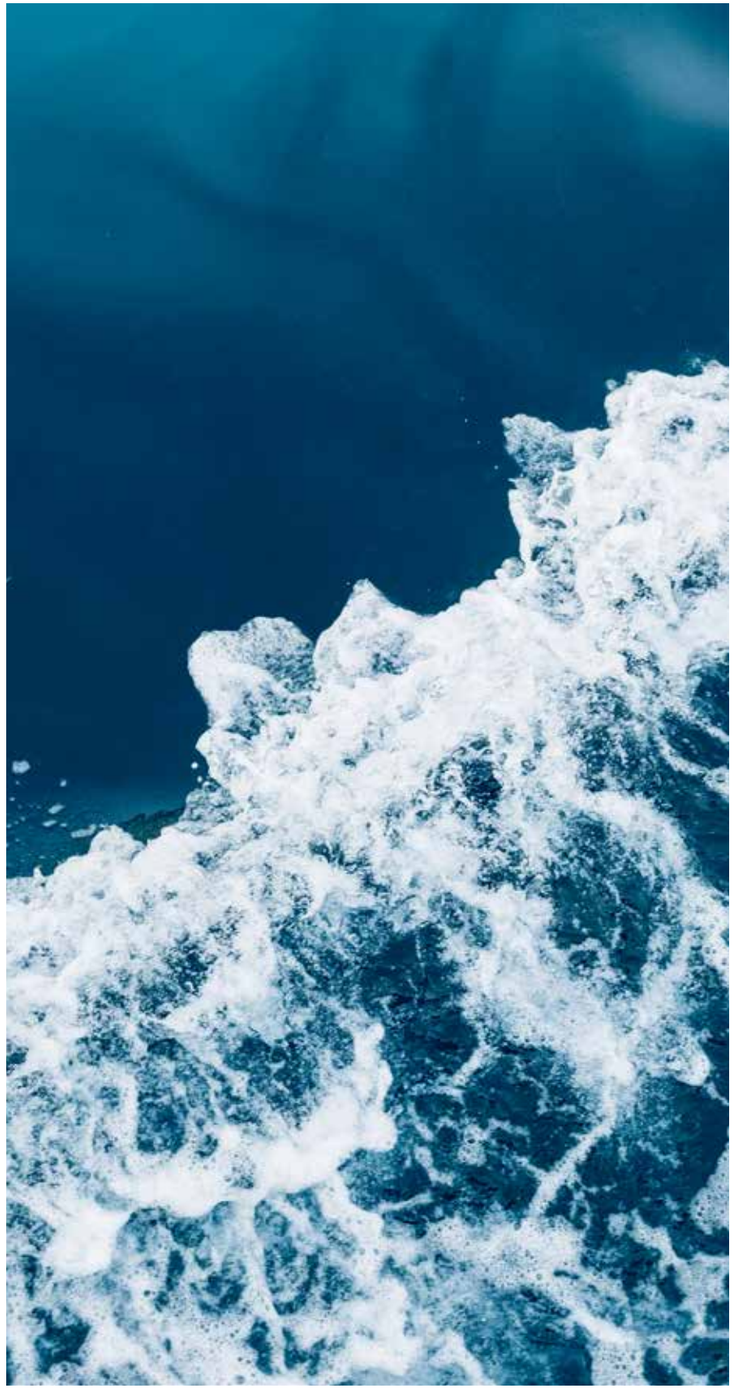
展望未来，持续开展跨学科研究有望揭示生命历史进程的新认知，同时推动革命性技术进步。通过伦理指引与 AI 分析能力的有机结合，我们将深化对地球生物多样性的理解，发展出支持生态与文明协同进化的技术体系。

©Tanyajoy / iStock / Getty

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).

第六章

地球与环境科学

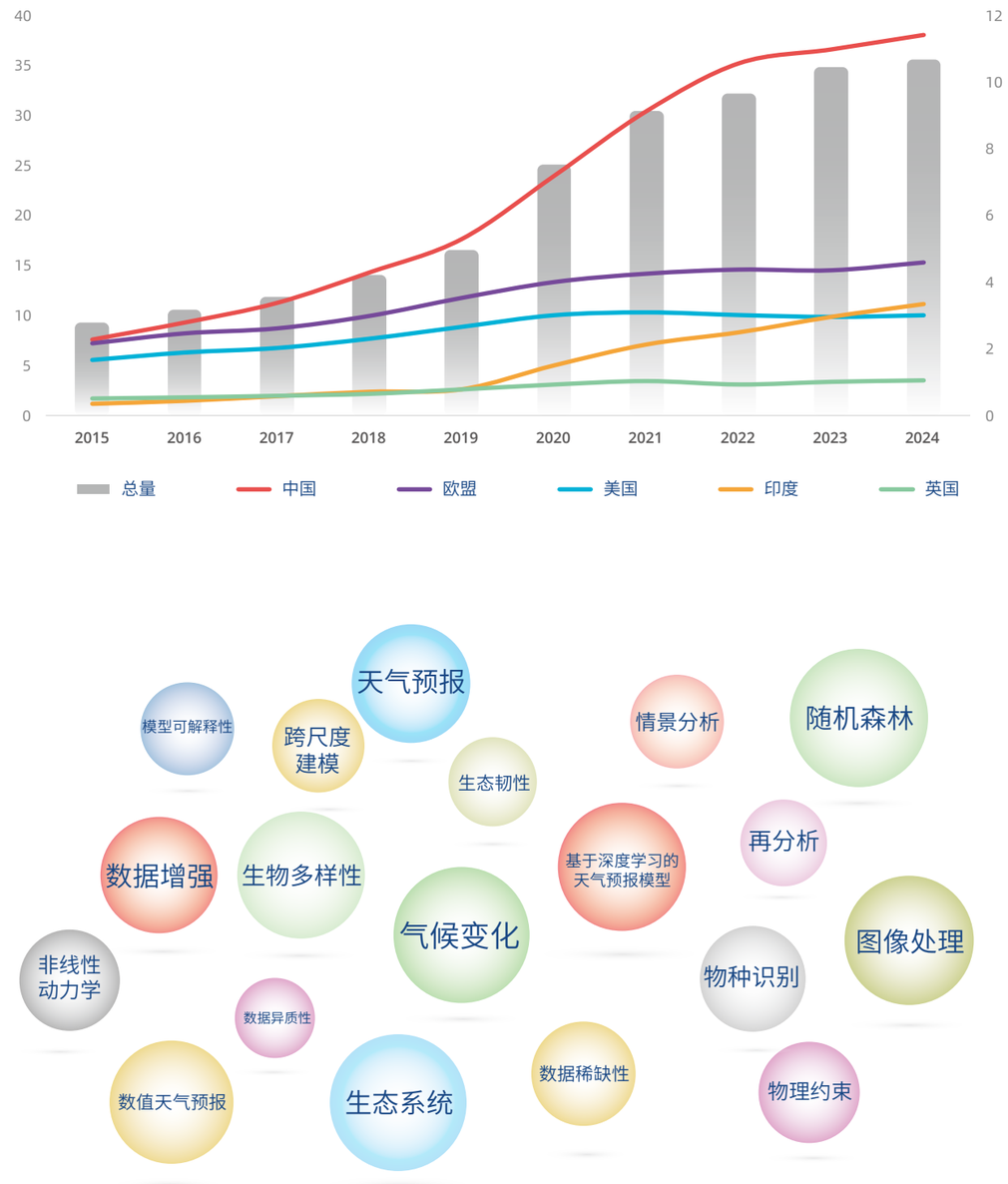


©Marija Siurtukova / Moment / Getty

AI 与地球环境科学

2015 至 2024 年间，地球与环境科学领域的 AI 出版物呈现四倍增长，学术出版物从 0.92 万篇跃升至 3.56 万篇（图 6）。中国贡献了全球近半数研究成果，大幅领先于欧盟和美国。同时，印度追赶态势明显，2024 年已超越美国。数据和关键词词云分析显示，气候变化、生态系统和生物多样性等议题最被关注。随机森林方法与图像处理技术应用广泛；多模态和端到端 AI 模型结合数值预测，极大改善了天气预报精度和效率；通过结合物理约束、数据增强技术和可解释性模型，AI 正在驱动地球和环境科学的范式转变。

图6 | 地球与环境科学领域AI出版物总量、国家趋势（单位：千篇）与关键词词云（2015-2024）



1. 大气科学

1.1 背景

气象预报在自然灾害防控、气象敏感型行业生产规划、人民生活和社会韧性建设等方面具有关键作用，其发展历程经历了从经验统计到物理建模的范式转变。20 世纪中叶，基于 ENIAC 计算机求解正压涡度方程的数值天气预报（NWP） 实现，标志着现代气象预报的开端。数值模式的预报精度在过去几十年间得到显著提升，这一进展被《自然》杂志评价为“静悄悄的革命”[1]。

数值模式包含两大核心模块：一是动力框架，用于求解大气演化的基本方程组[2]；二是物理参数化方案，用于近似描述次网格尺度的物理过程，如辐射、对流、云微物理、边界层以及陆面过程及其相互作用。然而，受限于观测数据缺乏、计算能力限制以及对物理过程理解受限，这些参数化方案不可避免地引入了不确定性，成为制约预报精度的瓶颈之一 。

值得注意的是，尽管计算能力取得了显著进步，但数值模式的预报精度并没有出现相应的显著提升。这一现象除了模式本身以及初值的影响外，也主要源于对经验性参数化方案的持续依赖，以及现代超级计算机上并行化所带来的计算复杂度激增[3]。在此背景下，人工智能（AI）技术为突破这一困境提供了新的范式革新机遇。

1.2 最新进展

近年来，基于人工智能的数据驱动模型在气象预报领域取得了突破性进展。自 2022 年起，多项前沿研究表明[4-8]，人工智能模型在中短期预报精度上已可比肩甚至超越国际上最先进的欧洲中期天气预报中心（ECMWF）的高分辨率确定性预报（HRES）。相较传统数值天气预报模式，人工智能数据驱动模型的显著优势在于其计算效率，仅需单块 GPU 即可在数秒钟内完成预报任务 。

在业务化应用方面，全球主要气象机构已加速推进人工智能气象大模型的业务部署。例如，中国气象局（CMA）于 2024 年 6 月 18 日正式发布了三款人工智能气象预报模型，分别为“风清”全球中期预报、“风雷”短临预 报和“风顺”次季节 - 季节预报，标志着中国在人工智能气象预报领域取得了重大突破。无独有偶，ECMWF 也于 2025 年 2 月 25 日将其人工智能预报系统（AIFS）投入业务运行，与传统数值模式形成协同互补的预报体系。

1.3 前沿科学问题和突破路径

AI 天气预报模型虽取得突破性进展，仍面临若干关键科学挑战，亟需跨学科协同攻关。

1.3.1 极端天气预测能力薄弱

现有模型普遍采用均方误差或平均绝对误差等损失函数， 虽优化了整体精度，却导致预报结果随预报时长增加而趋于平滑[9]。虽然自回归多步损失函数可缓解误差累积，但仍难以解决极端天气 (如强降水、大风等) 预报中细节丢失的问题。在气候变化加剧了极端事件的背景下，这一局限更为凸显。

突破路径：极端天气预测优化

概率生成模型[10]结合检索增强生成（RAG）技术， 通过历史相似案例挖掘 ，有望提升极端天气预报技巧。

1.3.2 过度依赖再分析资料

当前模型高度依赖 ERA5 再分析资料[11]， 但其降水数据与观测存在显著差异[12]，且热带气旋强度较 IBTrACS 数据存在系统性低估。这种单一数据依赖 严重制约模型在极端场景下的泛化能力。

突破路径：多源数据融合应用

高分辨率数值模式可生成合成数据[3]。例如，HR-Extreme[13] 等。通过数值模式降尺度和同化区域观测，构建区域极端事件数据集，弥补再分析数据对极端天气低估的局限性。

1.3.3 物理约束机制缺失

当前 AI 天气预报模型普遍缺乏物理约束， 可能导致出现违反物理规律的结果，如负湿度值[14]，或难以表征蝴蝶效应[15] 和地转平衡[9]。

突破路径：物理约束融合

物理和 AI 的混合建模框架（如 NeuralGCM 模型[16]） 通过可微分的动力框架求解离散化的控制方程。

1.3.4 时空分辨率受限

主流模型 0.25 的分辨率难以满足业务需求，而提升分辨率面临双重障碍： 1）高分辨率全球分析资料（如 ECMWF HRES 0.09。）样本量远小于 ERA5；其二，计算复杂度的增长。时间分辨率方面，6 小时间隔预报难以满足很多应用需求。

突破路径：时空分辨率提升

WRF 等区域模式可提供高分辨率的训练数据。时间分辨率方面， Pangu-Weather[5] 采用了分层时间聚合策略来实现 1 小时预报。

1.3.5 单一大气建模局限

现有模型多仅基于大气数据，而忽略海 - 气 - 陆耦合作用。极端天气事件（如热带气旋、洪水和热浪等）常由多圈层相互作用驱动，单一大气模型难以全面捕捉其机制。

突破路径：地球系统建模

通过构建耦合大气 - 海洋 - 陆地模型，可更准确表征海温、陆面过程对极端事件的影响。

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. 环境科学

2.1 背景

随着全球气候变化、资源紧缺、环境污染和生态破坏等问题日益严峻，环境科学在揭示自然规律、制定治理策略和保障可持续发展方面的起到重要作用，为应对环境污染和生态退化等全球性挑战提供了理论依据。借助机器学习和大语言模型等手段，人工智能（AI）能够实现遥感图像、实地观测和实验室数据等多源数据的融合，提高环境因子的时空精度，并实现对换环境因子变化的快速捕捉与精准预测，为环境科学带来深刻变革。AI 技术在构建数据驱动的多尺度、多过程耦合的环境模型、开展复杂系统模拟及不确定性量化等方面也展现出巨大潜力。未来 AI 技术将继续推动环境监测、治理和管理的智能化，为实现环境保护和可持续发展提供坚实支持。

2.2 最新进展

AI技术在环境科学领域取得了显著进展，应用范围不断扩大。

首先，基于卫星遥感、地面观测、地理信息和统计资料的多源数据融合技术极大地提高了环境因子时空特征变化分析的精度、环境监测与预警的时效性与准确性 [1,2]。AI 驱动的自动化处理系统能够迅速预估地震、野火、洪水、火山喷发等灾害的爆发和扩散，大幅提升灾害响应速度 [3,4]。自动化环境监测网络结合机器学习算法，实现了对环境污染、生物多样性变化、洪涝灾害的实时高效监控，有效降低了人工成本并提升了数据质量 [5]。

其次，AI 在环境系统建模方面的应用迅速拓展。深度学习与图神经网络的广泛应用，使包含大气、海洋、陆地等复杂地球系统模型、多尺度气候过程的模拟精度显著提高 [6]，提高了对生态系统、气候系统和社会经济系统的非线性关系和内在机理的理解，为环境政策制定及其效果评估提供了坚实的科学依据 [7-10]。同时，跨学科合作进一步促进了 AI 在灾害预警、物种保护、城市可持续发展规划等领域的落地应用。

此外，大语言模型（LLM）如 ChatGPT 和 DeepSeek 在环境科学研究领域的辅助作用也日益凸显 [11]。LLM 能够快速分析和处理海量环境数据，帮助研究人员从中提取有价值的信息，提升数据分析的效率和准确性。LLM 还可以辅助科研论文的高效撰写，生成高质量的文本内容，并为环境政策的制定与评估提供智能化决策支持。通过自然语言处理技术，LLM 促进了环境科学研究中跨领域的信息共享和知识传播，加速了科研进程。

2.3 前沿科学问题和突破路径

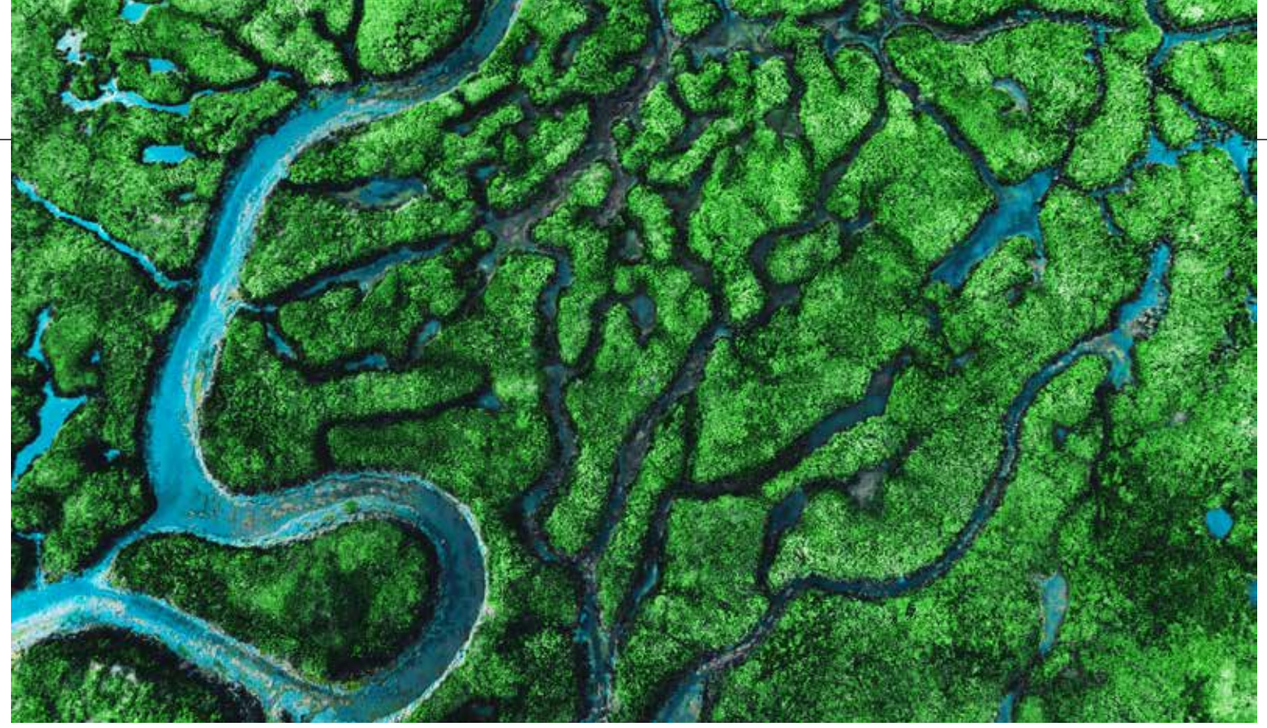
AI 与环境科学的交叉融合正在推动环境研究进入新阶段，核心方向体现在数据整合、机理融合和智能决策三个层面。

在复杂环境系统建模领域，如何整合卫星遥感、地面传感器、社会经济等来源各异的数据仍是一大挑战。数据在时空精度、取样标准等层面存在的显著差异，使得不同来源数据的时空对照，环境知识图谱的构建，及物理定律对神经网络的约束还存在技术差距。

环境临界点的预测则面临非线性系统突变的科学难题。多因素复杂交互使得传统模型难以捕捉其前兆信号，而 AI 技术为了寻求整体误差最小而容易忽略对极端事件和环境临界点的预测。

生态保护与经济发展的矛盾急需智能决策技术的创新，多智能体强化学习框架有效性及其有机整合是能否真正改变政策制定模式的关键 [12]。

AI 技术突破的关键在于环境数据的标准化和开放共享、环境机理与 AI 架构的深度融合、自然及社会经济过程的综合考虑。研究者应致力于制定统一的数据采集、处理和存储标准，建立跨机构、跨领域的开放数据平台，使来自卫星、地面传感器和无人机的实时数据、来自人类活动的统计数据和来自实验室的理论数据等多种来源的数据能够经过智能预处理，去除噪音和异常值。通过引入注意力机制、局部可解释模型和敏感性分析工具，研究者可以揭示模型内部的信息处理过程，同时结合环境科学、生态学原理和经济社会原理，开发出物理引导的深度学习模型，使得 AI 预测结果不仅准确，还具有科学可解释性。智能决策层面的突破需要实现多源数据、传统数值模型、数据驱动的模型及大语言模型的有机整合，从而实现问题、研究、决策和评估的一体化。



©Artur Debat / Moment / Getty

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. 生态科学

3.1 背景

生态科学旨在解析生态系统结构、功能与动态规律，对解决全球生物多样性丧失、气候变化等挑战意义重大。这些危机是复杂系统扰动引发的非线性动态失衡，难以预测。人工智能的突破为生态研究带来革新，其工具可量化和观测传统方法难以捕捉的生态现象，构建更精准的预测模型。

传统生态建模受系统复杂性限制，而 AI 深度介入正突破这一瓶颈。机器学习、复杂系统仿真等 AI 方法为生态建模提供新范式，推动学科双向创新。生态科学的复杂问题助力 AI 算法优化，AI 的算力优势加速生态规律发现。这种协同进化机制，为生态科学与人工智能的深度融合奠定方法论基础。

3.2 最新进展

3.2.1 物种分布与生态关系研究

通过 AI 算法，研究人员能够分析大量的生态数据，揭示景观模式与物种分布之间的复杂关系，这有助于更准确地了解物种的生存环境需求，为生物多样性保护提供科学依据。

3.2.2 多源数据融合监测

利用 AI 技术，能够将光学数据和雷达数据等多源数据进行融合，生成高分辨率的土地覆盖变化地图。可准确识别城市扩张和农业休耕草地的分布区域，以及森林损失和增加的趋势，为土地资源管理提供了实时、准确的数据支持。

3.2.3 生态过程模拟实验

借助深度学习中的长短期记忆网络（LSTM）等模型，可对生态过程进行模拟实验。如在模拟洪水对农业生态服务的长期影响时，通过强化学习建模结合多层感知器（MLP）神经网络和马尔可夫链分析，预测了城市化扩张和绿色基础设施部署的变化，为灾后土地管理提供了科学的决策依据。

3.2.4 知识引导的机器学习（KGML）

针对传统深度学习算法数据驱动、缺乏先验知识的问题，KGML 将科学知识注入机器学习算法的基础架构中，使模型能够做出更符合物理规律的预测 [1]。

3.2.5 深度学习与机理模型结合

将深度学习模型与传统的生态机理模型相结合，发挥两者的优势。深度学习模型能够处理复杂的非线性关系，而机理模型则具有明确的物理意义和解释性。通过这种结合，能够在提高模型预测性能的同时，增强对生态系统过程的理解和解释能力。

3.3 前沿科学问题和突破路径

3.3.1 模型的可解释性问题

许多 AI 模型，如深度学习模型，通常被视为 “黑箱”，难以解释其内部的工作机制 [2]。这使得研究人员在应用这些模型时，难以确定模型的预测是否合理，以及如何根据模型结果进行有效的生态管理决策。

突破路径：结合领域知识。构建生态系统模型时，融入生态系统的基本原理和规律，如能量流动、物质循环等，使模型能够基于这些知识进行推理和预测，从而提高模型的可解释性。

3.3.2 数据质量与数据稀缺问题

在生态科学领域，数据往往存在质量参差不齐、数据缺失、噪声干扰等问题，某些生态数据的获取难度较大，导致数据稀缺，使得模型在训练时缺乏足够的样本，从而影响模型的泛化能力和可靠性。

突破路径：多源数据融合。通过多源数据的互补，可以减少数据缺失和噪声干扰的影响，为 AI 模型提供更丰富、可靠的数据支持。

3.3.3 跨尺度和跨区域的模型泛化问题

目前的 AI 模型在跨尺度和跨区域的泛化能力方面存在不足，难以将在一个尺度或区域上训练的模型直接应用到其他尺度或区域，这限制了模型的应用范围和有效性 [3]。

突破路径：数据增强与迁移学习。在图像识别领域训练的模型，可通过迁移学习应用于生态遥感图像的分类和分析，从而在数据有限的情况下提高模型的性能 [3]。

综上所述，未来的研究应聚焦于解决数据稀缺、模型可解释性和跨尺度建模等前沿问题，推动 AI 与生态科学的深度融合。生态系统的复杂性和韧性为 AI 技术的发展提供了新的灵感，未来应进一步加强两个领域的交叉研究，推动 AI 与生态科学的共同进步 [4]。

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)

第七章

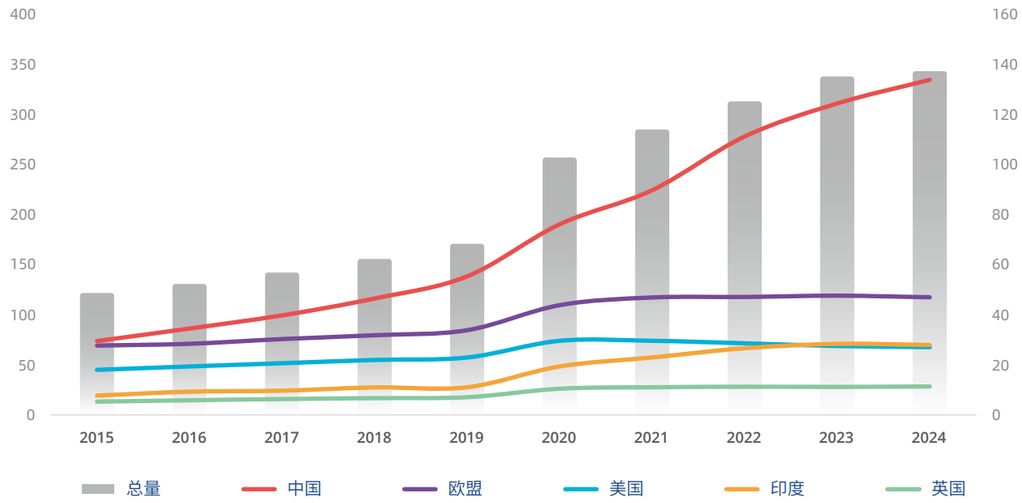
工程科学

©Xuanyu Han / Moment / Getty

AI 与工程科学

2015 年至 2024 年间，全球工程科学领域的 AI 出版物实现了近三倍的增长，相关出版物从 12.16 万篇增加到 34.30 万篇，中国出版物总量遥遥领先，远超欧美；而印度快速追赶，2024 年已超过美国（图 7）。数据和关键词词云分析显示，无线通信、网络优化、半导体设计和遥感等领域主题最被关注。大语言模型、联邦学习等 AI 方法已成为工程领域的重要工具；语义通信、边缘智能和模型分割等前沿技术将 AI 嵌入到下一代基础设施的骨干中，从而实现生态可持续性与系统韧性的深度耦合。

图7 | 工程科学领域AI总量、国家趋势（单位：千篇）与关键词词云（2015-2024）



1. 通信

1.1 背景

随着信息技术的飞速发展，人工智能（AI）与通信技术的融合正成为推动现代通信网络革新的核心动力。在第六代移动网络（6G）和第六代固定网络（F6G）的架构中，AI 被确立为核心使能技术 [1,2]，由众多业内公司联合成立的 AI-RAN 联盟 [3] 也将智能接入网作为重点发展方向。这一融合的意义深远：在技术层面，AI 使网络具备自我优化能力，提升可靠性、低延迟和能效；在产业层面，催生新的商业模式和应用场景，推动经济增长；在社会层面，支持智慧城市、智能交通和远程医疗等服务，显著提升生活质量。总之，AI 与通信技术的融合不仅是通信发展的必然趋势，也是实现未来通信愿景的重要途径。

1.2 最近进展

人工智能技术迅速发展，并正深刻变革通信领域的理论、实验手段和建模方法。从传统以香农信息论为核心的通信模式，到以语义通信为代表的新范式，AI 正驱动着通信学科的重大转变。语义熵、语义互信息和语义信道容量等数学模型的建立，有效拓展了经典信息论的边界，为设计高效、低延时的通信系统提供了全新理论依据。深度学习、多模态数据处理与智能决策实现了容量、覆盖和能效等核心指标提升十倍以上，打破了传统通信传输瓶颈。

与此同时，3GPP 等国际标准化组织正积极推动 AI 在无线网络中的应用。利用 AI 模型进行波束管理、信道预测及网络调度，不仅能够大幅提升系统响应速度和稳定性，还能为未来 6G 网络构建提供强有力的支撑。此外，分布式 AI 模型传输、智能控制和自适应决策等新技术，有望实现网络资源的最优配置和能耗的大幅降低。

此外，与通信技术跨学科融合正推动新型 AI 硬件加速器和算法平台的发展。新一代 AI 加速器与专用芯片借助超宽带互连网络，将使复杂神经网络的实时训练和推理成为可能，从而进一步支撑通信网络智能化水平的提高。

总体来看，AI 技术驱动的通信变革正朝着高效、智能、低耗的方向迈进。这一进程不仅为突破传统通信极限提供了全新解决方案，也预示着未来智能通信网络的无限可能。

1.3 前沿科学问题和突破途径

1.3.1 AI 赋能通信

尽管 AI 赋能通信为系统性能和效率提升提供了巨大潜力，然而深度神经网络对训练数据的依赖性较高。AI 赋能通信的一个重要科学问题是：数据依赖性和可解释性的平衡机制。针对该前沿问题，可以从数据、算法、系统和硬件等多个层面开展技术攻关。通过加入物理模型增强可解释性，降低计算延迟并防止陷入局部最优。上述关键技术的突破，有望实现从物理层到网络层全流程的协同优化，推动 6G 及下一代通信系统的稳定发展。

1.3.2 通信增强 AI 计算

通信系统作为算力传输的核心管道，正成为新型算力网络的关键支撑。在分布式 AI 训练中，通信增强 AI 计算的网络架构和协同机理成为一个关键的科学问题。针对该科学问题，在网络层研究软件定义网络的智能调度优化流量管理，提高通信效率；在物理层，光计算结合光互连技术被认为是突破传统电子计算功耗与速度瓶颈的重要方向，有望为 AI 计算提供更高效的通信基础。

1.3.3 原生智能的深度融合通信

AI 与通信的深度融合正从工具辅助阶段迈向系统原生阶段，使通信系统不再仅依赖传统优化手段，而是与智能技术深度耦合，形成“感知 - 传输 - 决策”闭环。这一变革催生了语义通信、脑机接口、通信大模型等新范式，但同时也带来了诸多理论问题。针对上述科学问题，需要在语义建模、计算架构、优化策略及资源调度等多个层面开展技术攻关，探索面向通信系统的大模型，将推动通信系统从传统优化模式向原生智能演进。

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).

©Yuichiro Chino / Moment / Getty

2. 遥感

2.1 背景

近年来人工智能技术快速发展，超大参数基础模型实现了颠覆性创新。遥感基础模型聚焦于多种遥感数据和图像解译任务的统一大模型逐步兴起，基本框架包括多模态遥感大数据、遥感基础模型以及地球观测类应用。遥感基础模型能够处理多源、多分辨率、多频段遥感图像，服务于场景分类、目标检测、跟踪、图像分割和变化检测等任务。融合 AI 技术与遥感科学，能够提升图像处理精度和多任务处理能力，满足日益复杂的遥感数据分析需求，推动遥感应用智能化发展。

2.2 最新进展

2.2.1 融合多时相多模态遥感数据的视觉大模型

融合多时相多模态遥感数据的视觉大模型，通过整合不同时间点和传感器的数据，实现地表特征动态监测。SatMAE [1] 基于 Transformer 架构对多光谱和 SAR 图像进行预训练，提升土地覆盖变化监测能力。空天・灵眸模型 [2] 通过掩膜自编码和 ViT 网络，构建了全球多时相数据集，验证了自监督预训练的潜力。

2.2.2 跨模态遥感图像解译的视觉语言大模型

跨模态遥感图像解译的视觉语言大模型，通过自然语言与视觉提示的交互，推动了遥感智能解译的变革。GeoChat[3] 实现了多功能视觉 - 语言模型，完成图像级、区域级、定位等复杂任务。EarthGPT [4] 为了缩小跨模态理解与视觉推理的差距，构建了百万级多模态数据集，支持光学、SAR、红外等多传感器图像的统一理解，可用于场景分类、图像字幕、目标检测等任务。

2.2.3 物理机理启发扩散模型的遥感图像生成

扩散模型 [5] 在遥感图像生成领域展现出巨大的潜力，通过逐步去噪生成高质量图像。将物理机理嵌入到扩散模型，引导模型在生成过程中充分考虑遥感图像的物理特性，DiffusionSat 将时空特性融入图像生成，优化空间一致性。HSIGene [6] 利用光谱特性优化扩散模型的注意力机制，从而有效捕捉空间频谱特征。

2.3 前沿科学问题和突破路径

2.3.1 探索遥感大数据与基础模型参数量之间的规模法则

多模态遥感数据融合与大模型规模法则及多源数据分布建模是当前研究热点，遥感视觉模型的“规模法则”[7] 尚不明确，亟待系统研究优化大模型构建。基于规范化构建的多时相多模态遥感大数据，建立基础模型的理论框架与规模法则的实证体系。通过系统性实验优化模型拓扑结构，并在不同参数规模下训练，探究数据规模与模型参数间的演化模式，揭示数据流形的本征维度与模型容量阈值的临界相变规律，并构建数学模型。

2.3.2 遥感数据自监督预训练算法及高效后训练技术研究

探索遥感数据高效融合架构与跨模态学习策略，通过自监督与监督学习训练技术，提升数据对齐与表征能力。遥感数据的自监督预训练与高效后训练技术，旨在降低对大量标注数据的依赖，提升模型泛化能力与训练效率。预训练阶段采用图像重建、生成和上下文预测等技术 [8] 构建高层次表征；后训练阶段聚焦于少量标注的高效微调，提高遥感任务精度与适应性。

2.3.3 遥感图像视觉任务的强化学习与奖励反馈机制研究

结合物理机制与数据驱动方法，结合强化学习与奖励机制，推动遥感智能化发展。通过引入强化学习与奖励机制 [9]，实现动态感知与自适应决策突破静态建模。构建虚实融合训练，结合物理辐射模型与真实卫星数据优化模型。设计多智能体协作解译，优化无人机与卫星监测，构建任务自适应奖励机制，融合地理时空先验，提升长周期任务探索效率。

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. 微电子

3.1 背景

随着器件尺寸逐渐接近物理极限，微电子技术面临着严峻挑战。首先，传统的硅基材料面临着越来越多的瓶颈，量子效应、热效应以及电流泄漏等问题逐步显现，导致器件性能特别是能效无法快速提升。其次，半导体工艺的复杂性不断增加，如何优化工艺参数以提高产量和良率成为亟待解决的问题。此外，电路设计的规模和复杂度日益增加，传统的设计方法在面对现代高性能、低功耗芯片需求时，设计效率和效果都亟待提高。

人工智能和机器学习技术的引入为微电子领域带来了新的机遇，能够高效处理大规模数据、优化复杂的设计和工艺流程，并通过数据驱动的方式发现潜在的规律和优化路径。这些技术能够有效解决传统方法无法应对的多维度、高复杂度的挑战。

3.2 最新进展

AI 技术的应用为材料科学提供了新的思路。生成式模型可以用来生成新的半导体材料 [1,2]。机器学习模型，能够预测包括稳定性、能带结构等关键性质 [3]。AI 模型还可以直接从目标性能出发，生成满足需求的材料 [3]。基于 AI 的高通量筛选，可以识别出具有潜力的候选材料，指导实验的设计和验证 [4]。

AI 在半导体器件建模和结构发现方面也展现出巨大潜力。神经网络和贝叶斯优化可以自动化器件模型的校准过程，提高仿真结果的准确性 [5]。在半导体器件建模中，生成模型能够从有限的实验数据中生成新的数据样本，从而扩展训练数据集，提升模型的训练效果 [6]。结合进化算法和机器学习，研究人员能够高效地搜索并优化半导体材料结构 [7]。

半导体制造涉及复杂工艺步骤和众多生产设备。虚拟计量技术利用机器学习模型预测生产过程中可能出现的质量问题 [8]。深度网络被用来分配任务到设备，提升了生产效率，减少生产周期 [9]。机器学习被应用于设备健康预测和故障检测，提前识别潜在问题，避免生产中断 [10]。贝叶斯优化等优化技术，能够通过最小化实验次数，找到最佳的工艺参数 [11]。

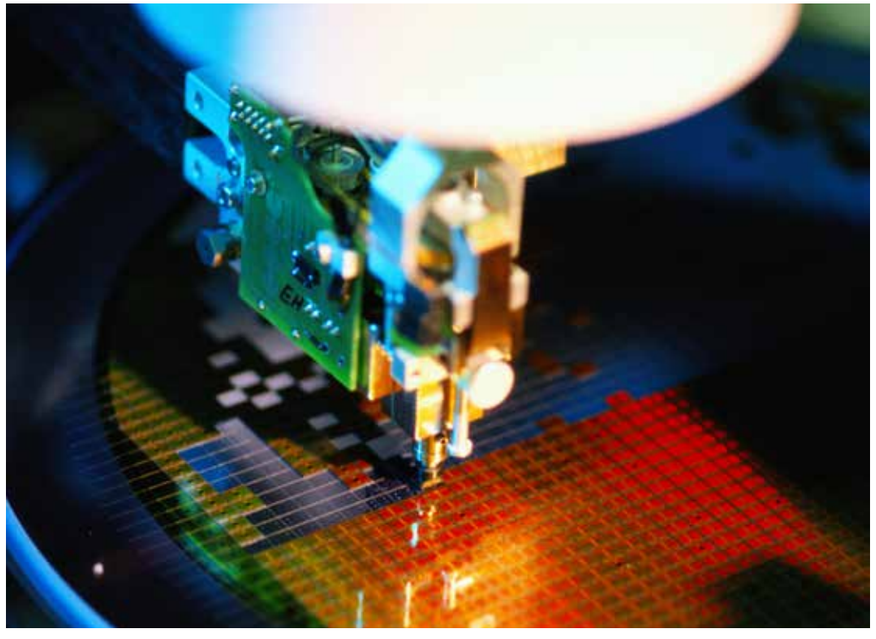
人工智能技术，在电路设计和 EDA 领域中应用广泛。神经网络技术被应用于射频 / 模拟电路的性能建模、数字集成电路设计中的拥塞情况估计 [12] 等。贝叶斯优化和强化学习技术，被应用于如模拟电路优化 [13]、集成电路的布局等。大语言模型能够自动生成硬件描述语言代码，根据用户需求自动化设计流程 [14]，调整模拟电路元件参数、修改电路拓扑等。

3.3 前沿科学问题和突破路径

生成式方法、可解释性以及大语言模型应用是 AI 在微电子领域应用的前沿问题。生成式 AI（如 GAN、VAE、扩散模型等）可用于新材料和电路结构的生成，并借助 DFT、分子动力学等方法验证性能，有效提升设计效率。然而，其在可控性、稳定性和计算开销方面仍面临挑战。模型可解释性问题限制了 AI 在高可靠性场景中的应用。通过特征可视化、注意力机制、决策树等方法，可揭示模型的决策逻辑，辅助工程师理解关键影响因素。大语言模型在微电子智能设计中展现潜力，但需应对领域知识复杂、推理准确度要求高等难题。

现有生成式方法主要依赖数据驱动，而微电子领域对物理规律高度依赖。可以结合物理启发的神经网络来保证生成结果符合基本物理约束。提升可解释性方面，可融合可视化技术、符号 AI 与知识图谱，帮助工程师理解 AI 决策逻辑。大语言模型应用上，应结合知识图谱生成微调数据，提升其推理能力与准确性。通过检索增强生成与形式化验证，确保 LLM 在复杂设计任务中的可靠性。同时，构建多智能体系统，实现材料、器件、工艺、电路等环节的全面智能化。

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. 空间信息

4.1 背景

随着 6G 网络和低轨卫星（LEO）星座的快速发展，通信系统正迈向 "AI 原生 " 与 " 空天地一体化通信网络 " 的全新阶段 [1-3]。6G 通过超低时延、超高带宽和海量连接重构底层架构，而星地协同网络融合近地轨道卫星、高空平台与地面基站，形成全球立体化通信基础设施。在此背景下，面向通信系统的大模型成为核心使能技术，支持无线信道预测、网络资源调度和跨域信号处理等复杂功能。然而，传统大模型在通信系统中面临算力 - 效能失衡、动态适配性不足、隐私与成本约束等挑战。边缘智能通过模型推理下沉至网络边缘（如基站、车载终端），结合轻量化压缩技术降低时延；星地协同计算则通过“在轨计算 - 星间组网 - 地面协同”架构优化分布式部署。边缘分割学习与星地拆分推理作为两大关键方向，通过大模型的分割与协同推理，突破资源约束、隐私保护及实时性瓶颈，为自动驾驶、智慧医疗、灾害应急等场景提供颠覆性解决方案 [4]。

4.2 最新进展

4.2.1 无线通信系统边缘分割学习

随着 6G 网络向“AI 原生”演进，边缘分割学习成为无线通信系统智能化的前沿方向 [5-7]。核心是将大模型动态分割为轻量化客户端子模型与服务器端计算子模型，通过梯度智能聚合与资源联合优化，突破边缘设备算力瓶颈。最新进展表明 [7]，动态梯度聚合筛选关键特征降低计算与通信负载，结合可调参数和深度强化学习预测最优分割点，实现自适应调度。联合资源优化策略针对设备异构性，采用智能子信道分配、动态功率调控及模型切割决策，显著降低训练延迟并提升资源利用率。

4.2.2 星地通信系统拆分推理

随着低轨卫星星座规模化部署，星地拆分推理技术在灾害预警、军事态势感知及气候建模等场景展现潜力 [8]。其核心将大模型按功能层级拆分，分布式部署于星端与地面站，优化计算与传输效率。最新进展显示 [9-11]，基于 Transformer 架构的模块化分割，将模型划分为星端浅层与地面深层子模型，动态任务调度结合轨道位置与链路质量调整策略。星地协同通信协议引入特征缓存与伪同步更新，配合稀疏编码与量化压缩减少数据传输量并保持精度。支持 Llama 3、DeepSeek R1 等开源架构的工具链，集成带宽模拟器与能耗评估模块，加速仿真到部署流程。

4.3 前沿科学问题和突破路径

4.3.1 算力 - 通信耦合与异构性

大规模分布式任务导致计算负载激增，且设备算力差异使分割策略难优化。为此，可开发多智能体强化学习实现实时动态调整分割策略，并结合星间组网提升整体算力。

4.3.2 动态环境适配

星地链路及无线环境的波动影响实时推理的稳定性，为解决此问题，可设计高效容错通信协议，利用特征缓存与数据压缩技术确保推理过程的连续性。

4.3.3 隐私 - 效率权衡

梯度聚合与数据压缩会导致信息损失，同时遥感数据下传存在隐私泄露风险。可通过引入差分隐私与联邦学习优化聚合过程，并搭配轻量化加密技术，提升隐私保护与效率平衡。

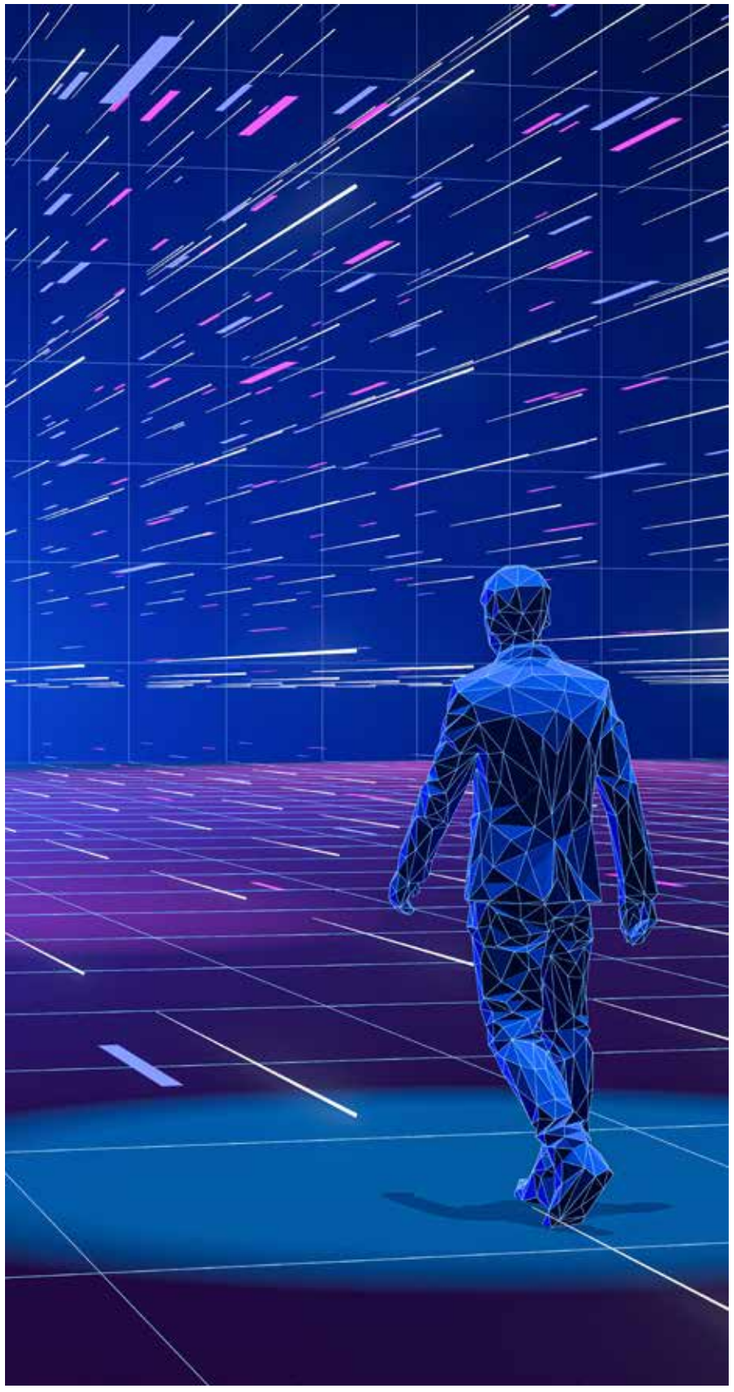
4.3.4 多维资源优化复杂性

信道分配、功率调控与模型分割的联合优化构成非凸问题，难以满足毫秒级延迟需求。可建立分割学习与拆分推理的收敛性数学模型，量化资源约束对性能的影响，为优化提供理论支撑。

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. 李力，戴阳利 . “ 新基建 ” 背景下卫星互联网发展的机遇和风险 . *卫星应用* **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).

第八章

人文社会科学

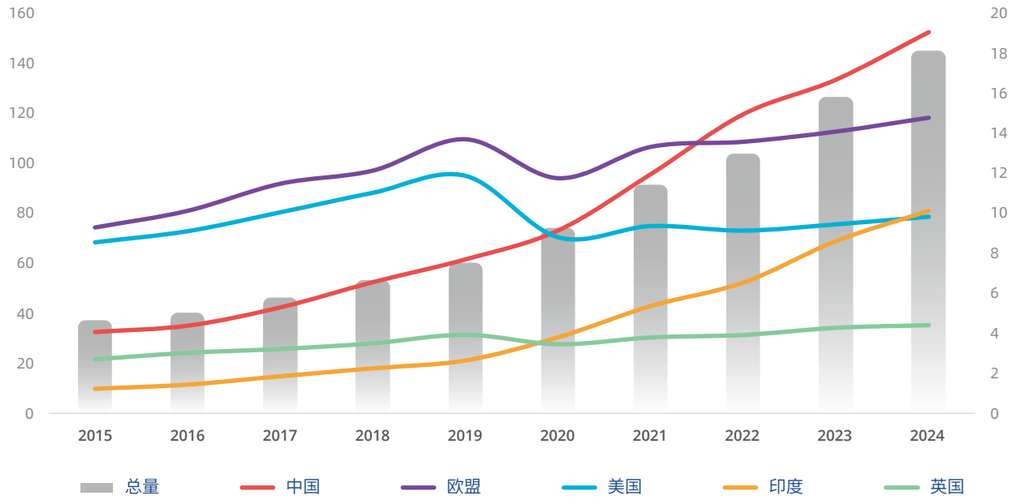


©marian / Moment / Getty

AI 与人文社科

2015 至 2024 年间，人文社科领域的 AI 研究实现了近四倍的增长，相关出版物从 3.71 万篇增加到 14.48 万篇（图 8）。中国于 2022 年超越欧盟，成为该领域出版物的最大产出国，印度也快速追赶，于 2024 年超越美国。数据分析和关键词词云显示，研究主要聚焦于 AI 透明性与伦理治理等，通过隐私保护、减少偏见和价值对齐的框架来应对 AI 的社会风险。从 AI 方法来看，基于多主体建模、多模态大模型和复杂系统的研究工具，正在重新建构解析复杂经济和社会系统，解构文化现象，推动人机交互新范式，这将对未来数字经济和社会的可持续发展产生深远影响。

图8｜人文社会科学领域AI总量、国家趋势（单位：千篇）与关键词词云（2015-2024）



1. 社会科学

1.1 背景

传统社会科学研究范式经历了四个发展阶段：从经验驱动的定性研究；到理论驱动的定量研究；再到机理驱动的仿真模拟，以及数据驱动的大数据分析。AI 技术催生了“数据和机理双驱动”新范式，能够全面革新人文社科研究范式，提高研究效率，更提升了研究的广度和深度。

1.2 最新进展

AI 给人文社科研究带来了三方面的革命。一是科学研究对象的革命，二是模型构建复杂性的革命，三是预测能力的革命。

1.2.1 数据挖掘和处理能力明显提升

生成式 AI 在零样本或少样本的情况下能够有效执行特定的文本分类任务。研究显示，生成式 AI 在零样本环境下识别虚假信息、立场和情绪分类等任务中，准确率很高 [1,2]。

1.2.2 复杂系统建模能力大幅增强

基于多主体建模（Agent Based Modeling，ABM）的发展，给复杂社会系统仿真带来了突破。ABM 可以刻画微观主体的有限理性，以及非均衡、非显性的交互模式，极大丰富了经济社会系统模拟的复杂性。这些特征使其在政治事件、金融市场等复杂性、异质性和易变性较强的场景下，尤为适用 [3,4]。

1.2.3 预测与决策能力全面升级

计算社会科学是基于大规模数据和高通量计算，对个体及群体行为进行推演和计算的科学，其核心目标是指导更科学的决策。

在个体决策方面，AI 算法与随机优化的结合改变了“先预测后优化”的传统范式，提高优化效果和效率 [5,6]，适应更加动态、复杂的运营环境下 [7]。公共决策方面，基于 ABM 推演多智能体复杂行为，捕捉特定环境下的交互及由此涌现出的社会结果。

1.3 前沿科学问题和突破路径

1.3.1 模型的复杂性与机理融合

AI 技术虽然在预测与分析能力上表现出色，但其解释性不足和机理融合不够，对社会科学中强调的理论建构形成了巨大挑战。主要攻关方向包括提升 LLM 与 ABM 的机理融合能力，通过加强跨学科的理论与算法合作优化模型设计，以及引入多尺度建模方法实现宏微观贯通。这些方法有助于捕捉社会现象的动态演变过程。

1.3.2 机理的不确定性与博弈机制的挑战

社会科学中涉及的人类行为往往具有高度的随机性和非线性，对传统模型提出了巨大挑战，亟待在 AI 模型中引入不确定性量化，并将之与博弈机制融合。通过强化学习等技术，AI 能够自适应学习社会个体的决策模式，从而实现对复杂社会现象的动态预测和政策模拟。

1.3.3 大模型的偏见问题

生成式 AI 作为社会研究的新方法富有潜力，但也存在样本有偏的问题。大模型更容易代表教育 / 收入水平高、政治观念激进、以及西方国家群体的民意，而对政治、宗教、社会经济地位处于边缘的群体的观念呈现不足 [8,9]。不能简单认为“硅基样本”可以替代人类，对其能力和局限要给予充分的实证研究和持续关注。未来需要通过合适的训练方法，如增加多元化数据来源和设计去偏见算法，以逐步消除语言模型的偏见，推动生成式 AI 更加公平和包容。

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).

©Andriy Onufriyenko / Moment / Getty

2. 人文科学

2.1 背景

AI 技术与遥感成像、历史地理大数据、虚拟现实等技术结合，给历史现实的发现、分析和重现带来全新的模式。推动了从墓葬识别、文物识别、文化遗产保护到考古遗址数字化重建的全面创新；深化分析时空数据、重建人类文明变迁的多模态时空格局；加速破译未知文字、分析语义变化。

2.2 最新进展

2.2.1 AI 与遥感技术结合加速遗址发现和发掘

传统的考古受地形、植被和人类活动等因素的限制，存在可访问性差、记录不均衡、主观性强等问题。遥感技术与 AI 的融合，为大规模、高精度的遗址检测提供了新的可能。利用人工智能算法，可以处理海量的遥感数据，从中自动识别潜在的考古遗址。[1,2]AI 与遥感结合在多维度数据挖掘、历史图像分析以及复杂地理结构识别中的独特优势，为大规模、高精度的遗址发现和发掘提供了可能。

2.2.2 多模态 AI 技术对人类与文明演进过程的深度解析

在历史地理领域，运用 NLP 技术可自动化地从大量历史文献中提取关键信息。计算机视觉技术则可以分析历史地图和图像，自动识别地理特征和地标，生成详细的地理信息图层。通过多种数据源相结合，能够提供更为丰富和立体的历史地理信息。

语言文字与 AI 的结合不仅可用于甲骨缀合 [3]、文字识别 [4] 和考释 [5] 等，还可利用 LLM 智能体打造的“AI 古文字专家”，进行汉字教育及文化推广等。

2.2.3 生成式 AI 创新文化展示与传播

AI 与虚拟现实（VR）、增强现实（AR）等技术的结合，开辟了公众参与的新方式。数智城市和元宇宙技术通过精细的 3D 建模，将历史建筑、遗址场景在数字空间中完整重现，用户不仅能够“穿越”到过去的场景中体验古代文化，还能参与虚拟考古活动，为文化遗产的保存和传播带来更具互动性的方式。

2.3 前沿科学问题和突破路径

人文学科的数据来源广泛且极为复杂，来源的异质性和质量的不一，显著增加了 AI 模型训练的复杂性与难度。数据标准化与质量控制成为关键难题，亟待研究人员和技术专家制定统一规范和严格的质量评价标准，以确保数据的可靠性与有效性。

AI 在理解和解析复杂历史文化背景时常出现偏差，“意义壁垒”尤为明显。历史事件和文化现象往往具有高度的背景依赖性，许多信息背后的含义隐含于特定的文化、历史语境之中，这种微妙的内涵往往难以被机器完全理解或准确解读。因此，人机协作模式的重要性日益凸显，AI 应明确定位为辅助工具，帮助人类学者更高效地进行分析、解读和推断，而非取代人类专家的深度理解和判断。

面向未来，数字人文的发展应当聚焦于三大领域：人机协作、多学科融合与技术创

新。应加强人文学者与 AI 技术人员之间的协作，开发出既满足学科需求又充分发挥 AI 潜能的智能工具。推动跨学科数据库建设，整合历史学、文学、哲学、社会学等多领域数据资源，建立统一且高质量的信息平台。在人才培养上需加大投入，培养一批兼具人文学术背景和 AI 技术能力的复合型人才。

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 伦理治理

3.1 背景

AI 对人类的影响可以从短、中、长期进行分析。从短期来看，各类智能工具有效提升了工作效率，方便了生活；从中期来看，智能系统改变了社会运行、价值结构；从长期来看，超级智能对人类的影响不仅限于表层的社会活动，更可能触及深层次的生存问题。

随着 AI 在医疗、自动驾驶和金融等高风险领域的应用，其潜在风险不断凸显，涉及隐私和歧视、道德判断差异、自主系统的责任真空 [1,2] 等。当下对 AI 伦理治理的探索主要聚焦于弱 AI 技术，关注公共政策、应用伦理以及 AI 提供的价值 [3]。但着眼更长期，通用 AI、强 AI，尤其是超级智能可能带来的影响是将来 AI 伦理与治理研究的关键方向。

3.2 最新进展

3.2.1 AI 透明性与可解释性问题

透明性和可解释性是 AI 伦理研究中的核心议题之一。学者们提出了多种方法来提高 AI 系统的可解释性。比如通过为每个输入特征分配一个重要性分数，来解释复杂模型的输出 [4]。此方法广泛应用于金融、医疗等领域，以帮助人们理解 AI 系统的决策机制。

3.2.2 隐私问题

近年来，AI 在处理个人数据、个体生物信息分析等方面展现了强大的能力，但也引发了隐私侵害的担忧。研究者们提出了多种保护措施和技术手段，例如联邦学习、差分隐私等 [5,6]。

3.2.3 偏见问题

AI 系统可能会延续社会偏见或引入不公平 [7]。有学者认为，可以通过设计、训练和部署 AI 模型中的道德和法律原则，来减轻 AI 系统的偏见，从而在充分发挥 AI 技术潜力的同时确保社会利益 [8]。

3.2.4 失控问题

随着通用人工智能（AGI）、甚至超级人工智能（ASI）逐渐显示出现实迹象，超级智能体可以出现失控问题。比如面临关机问题进行自我复制从而越过人类红线或者由于追求目标的无限、自主的行动而演化出回形针问题。

3.3 前沿科学问题和突破路径

3.3.1 价值对齐问题

如何确保 AI 系统的目标与人类价值观和利益相一致，是通用 AI 和强 AI 伦理研究的关键议题。可以通过人类反馈来训练 AI 系统，使 AI 系统的行为逐渐符合人类的预期 [9]。此外，Anthropic 公司提出 AI 价值对齐的 3H 原则——有用（helpful）、诚实（honest）、无害（harmless），成为了诸多对齐研究的范例 [10]。未来，面向全球化的多元文化价值对齐研究有望成为下一个重点。

3.3.2 公平性与责任问题

AI 模型在训练过程中可能会学习到源数据中固有的偏见，从而表现出不公平的决策行为。偏见缓解技术通过在 AI 模型开发多个阶段进行干预，减少或消除系统中的偏见 [11]。公正性指标也成为评估 AI 系统公平性的重要工具 [12]。

随着 AI 逐渐参与到决策过程中，如自动驾驶汽车和自动化医疗诊断，明确责任划分变得尤为重要。Floridi 和 Cowls（2019）提出的“AI 伦理五原则”——善意（beneficence）、不伤害（non-maleficence）、自主性（autonomy）以及公平正义（justice），可解释性（explicability），明确了 AI 系统中责任性的核心地位。

3.3.3 失控与伦理应对路径

从“以人为本”“以人为中心”到“以人为目的”构建掌控超级智能的伦理原则。除了亚里士多德美德伦理可以提供洞见 [13]，奉献美德和智能契约伦理会成为可能的伦理应对路径。

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

第九章

展望与政策

©Andriy Onufriyenko / Moment / Getty

1. 未来挑战与研究方向

人工智能（AI）技术正以前所未有的深度和广度渗透到数学、物质科学、生命科学等重大科学领域，也推动这些领域中前沿科学方向和问题的突破。这些挑战既涉及 AI 基础理论的创新突破，也包含跨学科融合带来的范式变革，更需应对由此产生的伦理与社会治理难题。

1.1 AI 模型进化：从专用到通用

当前 AI 的核心挑战在于如何扩展模型能力的边界，以满足现实场景的多样性需求。深入理解 AI “黑箱”运行机制的关键是揭示神经网络内蕴的数学理论与强化学习的收敛机制，而下一代扩展定律和高效推理方法将决定模型性能的天花板。未来，基于群智涌现的多智能体协作和全模态统一建模，有望实现具备跨本体泛化能力的通用 AI，但需突破动态环境适配和多维资源优化等瓶颈。

1.2 超学科融合：重构科学研究新范式

AI 正打破学科间的壁垒，形成连接不同学科的超级纽带。例如，蛋白质智能设计与生成式材料模型结合，催生了具有生物响应性的智能材料；脑科学的调控难题正通过脑机接口和生物启发模型寻找突破口，同时也推动新一代类脑计算器件的快速发展。AI 构建的多尺度建模框架正建立从微观分子到宏观生态系统的统一理论，揭示跨学科领域的共性规律。这种超学科融合催生了“生物 - 信息 - 物质”等新兴交叉领域，推动科学从分析走向综合、从局部走向整体。

1.3 伦理与安全：构建 AI 的“刹车系统”

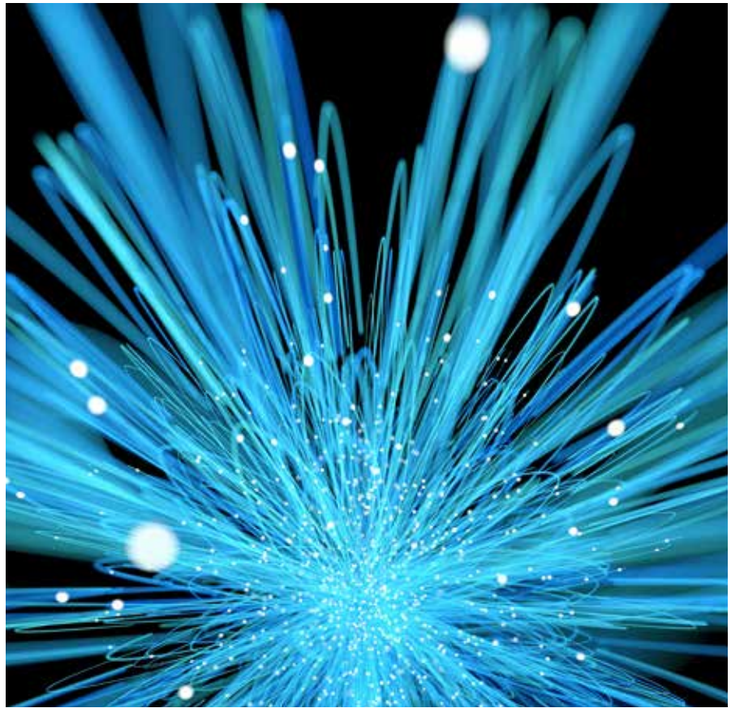
随着 AI 渗透到医疗、具身智能等关键领域，其安全治理问题日益凸显，主要体现在三个方面：在数据层面，隐私保护与数据共享的矛盾，需要新型加密计算范式的创新；在技术层面，模型可解释性与公平性的缺失，导致临床决策等应用场景存在信任危机；在价值层面，AI 价值观对齐和伦理框架构建，决定着技术进化方向。未来亟需建立动态风险评估体系，通过技术创新与制度设计的协同配合，为人工智能的安全可控发展提供坚实保障。

2. 政策框架

2.1 政策目标

随着 AI 技术的快速发展，AI 在科学研究中的应用潜力日益凸显。AI for Science（AI4S）旨在通过 AI 技术推动科学研究的创新与突破，提升科研效率，解决复杂科学问题。为此，制定全面的政策至关重要，以确保 AI 在科学研究中的有效应用，并最大化其社会和经济价值。

政策的核心目标是加速 AI 技术与科学研究的深度融合，促进跨学科协作，推动科学发现模式的变革。具体而言，我们希望通过 AI 提高科学数据分析的效率，优化实验设计，强化模拟与预测能力，以解决传统科



©Weiquan Lin / Moment / Getty

研方法难以应对的复杂问题。同时，政策还应旨在构建开放、透明、安全的科研生态，促进数据共享与知识交流，降低研究壁垒。此外，为保障 AI 技术的可持续发展，应加强对人才培养、伦理治理及法律规范的支持，确保 AI 在科学研究中的应用符合社会价值观与伦理准则，从而推动科技创新、产业升级与社会福祉的全面提升。

2.2 政策制定

为实现 AI4S 战略目标，建议政策围绕以下六大核心领域构建系统化实施方案：

2.2.1 数据共享

推动科研机构、政府与企业之间的数据互通，建立标准化科学数据共享平台，鼓励开放获取与跨学科数据整合。制定数据质量评估标准，确保数据的可靠性与可用性。

2.2.2 安全及隐私保护

强化科学数据的安全管理机制，实施严格的访问控制与加密技术。针对涉及个人隐私和伦理敏感的科学数据，建立合规审核机制，确保 AI 技术的应用符合伦理要求与法律法规。

2.2.3 算法开发

统筹推进基础算法研发与场景应用突破，构建开放协同的技术创新体系。推动关键算法开源共享与迭代优化，形成产学研用联动的技术攻关机制。

2.2.4 人才培养

完善 AI 与科学研究交叉学科的教育教学体系，支持高校开设 AI4S 通识与专业课程。推动产学研协同育人，设立“师生共创”的 AI4S 专项培训计划，培养具备跨学科能力的复合型拔尖创新人才。

2.2.5 资金支持

设立专项科研基金，支持 AI 在科学研究中的基础研究和应用探索。建立多层次的资助机制，为初创企业和科研团队提供长期稳定的资金支持，以促进创新成果的转化与落地。

2.2.6 法律伦理

制定 AI 在科学研究中的伦理指南，明确责任归属，确保 AI 技术的公平、公正应用。建立跨部门监管机制，确保 AI 技术的应用符合法律框架，并在社会监督下健康发展。

2.3 政策实施

2.3.1 执行机构

AI4S 政策的执行涉及多主体协同。政府部门负责政策制定与统筹规划；科研机构承担技术研发与应用示范，如实验室、高校等，通过组建跨学科团队开展前沿研究；企业则聚焦技术转化与产业化，如科技企业参与构建 AI4S 创新生态。

2.3.2 组织架构

构建多层次、协同化的组织架构。设立各级 AI4S 战略委员会，统筹协调各部门资源与政策；在地方层面，建立区域 AI4S 推进小组，负责落实政策与项目对接；同时，鼓励产学研用各方组建联盟或共同体，形成开放合作的创新网络。

2.3.3 资源配置

整合多渠道资源，保障 AI4S 发展。财政方面，设立专项基金支持基础研究与关键技术研发；科研资源上，推动算力、数据、模型等共享，人才资源方面，加强 AI4S 专业人才培养，设立相关学科与培训项目。

2.4 政策评估与调整机制

2.4.1 监督机制

建立多维度监督体系，政府部门负责政策执行的宏观监督，确保政策方向与目标一致；第三方评估机构对项目实施效果进行独立评估，提供客观反馈；同时，引入社会监督，通过公开透明的信息发布机制，接受公众监督。

2.4.2 反馈机制

构建动态反馈渠道，政策执行主体定期向监督机构报送进展情况；设立专家咨询委员会，收集科研人员、企业代表等各方意见，及时反馈问题与建议；利用大数据与人工智能技术，监测政策实施的实时数据，为反馈提供数据支撑。

2.4.3 评估与调整

定期开展政策评估，从技术进步、产业带动、社会影响等多维度衡量政策效果；根据评估结果与反馈信息，及时调整政策内容与资源配置，优化政策实施路径。

附录一：数据源

Dimensions 数据库：作为综合性科研信息资源库，Dimensions 汇集了超过 1.5 亿条文献数据、4200 万个数据集、790 万条科研经费数据、1.67 亿条专利数据、90 万条临床试验数据以及 250 万条政策文档数据。通过灵活整合和关联不同类型和不同层级的数据，Dimensions 数据库有效协助研究人员进行多维度研究。研究人员可以在国家 / 地区、城市 / 都市圈、机构、个人等层面上，结合文献出版、引用、基金、临床试验进行分析。研究人员也可对比分析不同机构或作者的研究主题、优势学科、合作网络等方面。

自然指数 （Nature Index）：自然指数提供简单、透明的指标以量化全球高质量研究。自然指数追踪 145 个高质量自然科学和健康科学期刊中的科研论文。期刊选择由科学家组成的独立小组选出。

附录二：数据使用和说明

1. AI 相关出版物

以 Dimensions 数据库中“人工智能”领域相关出版物（种子文章）为基础，通过构建模型，利用种子文章的标题和摘要嵌入，逐步迭代匹配语义相近的出版物，从而扩展数据范围。另外，额外补充了与 Dimensions 数据库匹配的中国计算机学会 A 类刊物（CCF-A）会议论文数据。

2. AI 核心和 AI4S 主题划分

2.1 根据 AI4S 白皮书需求，将 AI 相关出版物按照 FOR 学科一级学科分类进行相应主题映射：

2.2 因同一篇出版物可能被划分到多个 FOR 学科一级分类，如果该出版物属于同一白皮书主题划分，则在该主题下进行去重处理。

3. 统计国家 / 地区相关数据时，将未知国家 / 地区数据排除在外。

4. AI 前沿和 AI4S 主题下，领域科学问题与 AI 技术的呈现（即图 1.8 和图 2-8 词云部分）为示意图，非数据图，图例相对大小不代表实际数据。

FOR 学科一级分类	AI4S 白皮书主题划分
信息与计算科学 (Information and Computing Sciences)	AI 核心
地球科学 (Earth Sciences)	地球与环境科学
环境科学 (Environmental Sciences)	地球与环境科学
工程学 (Engineering)	工程科学
商业、管理、旅游与服务 (Commerce, Management, Tourism and Services)	人文与社会科学
经济学 (Economics)	人文与社会科学
教育学 (Education)	人文与社会科学
历史、遗产与考古学 (History, Heritage and Archaeology)	人文与社会科学
人类社会 (Human Society)	人文与社会科学
语言、传播与文化 (Language, Communication and Culture)	人文与社会科学
法律与法学研究 (Law and Legal Studies)	人文与社会科学
心理学 (Psychology)	人文与社会科学
生物学 (Biological Sciences)	生命科学
生物医学与临床科学 (Biomedical and Clinical Sciences)	生命科学
健康科学 (Health Sciences)	生命科学
数学 (Mathematical Sciences)	数学
化学 (Chemical Sciences)	物质科学
物理学 (Physical Sciences)	物质科学

备注或披露：

1. 数据更新时间：2025 年 4 月
2. 附录数据说明部分使用生成式 AI 润色文字

自然科研智讯（Nature Research Intelligence）

自然科研智讯致力为政策制定者、科研管理者、研究机构、资助机构及企业等提供深入的科研分析和科研全景概览。 通过整合 Dimensions、Nature Index、Crossref、 OpenAlex 等来源，关联分析科研文章、经费、专利、临床试验和政策文件等数据，灵活组合自然指数 (Nature Index)、自然引航 (Nature Navigator)、自然策略报告 (Nature Strategy Report)，自然科研智讯提供科研表现数据分析，洞察科研发展趋势，揭示研究潜力和机会， 支持战略决策制定。


復旦大學
SAIS 上海科学智能研究院
Shanghai Academy of AI for Science
nature
research intelligence