

# 科研智能发展报告

## (2025 年)

中国信息通信研究院人工智能研究所

中国人工智能产业发展联盟

2026年1月

---

## 版权声明

---

本报告版权属于中国信息通信研究院、中国人工智能产业发展联盟，并受法律保护。转载、摘编或利用其它方式使用本报告文字或者观点的，应注明“来源：中国信息通信研究院、中国人工智能产业发展联盟”。违反上述声明者，编者将追究其相关法律责任。

## 前 言

当今世界正经历新一轮科技革命与产业变革，科技创新正从“要素驱动”加速转向“知识驱动”。大量理论与实践表明，科学研究与技术创新并非经济活动的附属品，而是支撑长期增长与国家竞争力跃升的内生动力；研发投入强度的长期分化，正在把创新优势固化为结构性竞争力差距。在此背景下，人工智能与科研活动深度融合，推动科研范式从以人为主的线性流程，迈向数据—模型—计算—实验协同的闭环体系，科研智能由此成为全球科技竞争的新焦点。

面向这一战略赛道，各主要经济体纷纷出台专项政策与重大计划，通过“顶层战略牵引—算力与数据底座—组织化科研投入—场景任务牵引”联动布局，形成面向基础科学突破与产业研发转化的系统性支持框架。与此同时，科研智能关键技术也在快速演进：科研模型正从通用大模型走向面向科学知识表达、科学推理与科学对象表征的增强体系；科研智能体将“理解—规划—工具使用—环境交互”贯通为可迭代的科研工作流；自动化实验室则把算法决策与实验执行深度耦合，推动“干湿闭环”从概念走向工程化实践，上述技术共同推动了科研范式向“第五范式”的加速跃迁。在生物医药、新材料、半导体与先进制造等关键领域，一批代表性进展正在拓展对关键机理与规律的理解边界，并通过提升设计与验证效率，逐步改善产业研发长期面临的高成本、长周期与低成功率问题。

本报告旨在系统梳理科研智能的发展背景、政策举措、关键技术体系与典型应用，服务政府部门、科研机构与产业界把握趋势、识别

路径、凝聚共识，并为相关战略制定与工程落地提供参考。报告自 2025 年 9 月启动编制，综合采用文献研究、案例调研与专家访谈等方法，力求在宏观格局与关键细节之间取得平衡。需要指出，科研智能仍处在快速演化期，本报告相关研判以 2025 年底公开资料与专家观点为基础，期望以阶段性研究抛砖引玉，推动各方在实践中不断校准、迭代与完善。



# 目 录

一、 科研智能发展背景 .....	1
(一) 科研创新是经济增长的内生动力 .....	1
(二) 科研范式的历史演进与当代局限 .....	4
二、 科研智能发展历程 .....	7
(一) 概念及内涵 .....	7
(二) 发展历程 .....	8
三、 科研智能政策举措 .....	12
(一) 美国 .....	13
(二) 欧盟 .....	14
(三) 中国 .....	16
(四) 其他国家 .....	18
四、 科研智能关键技术 .....	20
(一) 科研数据 .....	20
(二) 科研计算 .....	23
(三) 科研模型 .....	25
(四) 科研智能体 .....	28
(五) 自动化实验室 .....	31
五、 科研智能典型应用 .....	33
(一) 驱动基础科学突破 .....	33
(二) 加速产业研发进程 .....	49
六、 发展挑战与展望 .....	64
(一) 发展挑战 .....	64
(二) 未来展望 .....	68

## 图 目 录

图 1 全球主要经济体研发强度示意图（2000 年至 2021 年） .....	3
图 2 科研智能关键技术示意图 .....	20
图 3 PANGAEA 从数据提交到发布使用的工作流程示意图 .....	22
图 4 谷歌 AI Co-Scientist 多智能体架构设计示意图 .....	31
图 5 利用分布式实验室发现新材料的流程示意图 .....	33
图 6 模块化机器人工作流程和启发式反应规划器示意图 .....	40
图 7 ECMWF AI 预报系统 AIFS 工作流程示意图 .....	44
图 8 AI 发现药物分子在临床试验中的成功率示意图 .....	53
图 9 生成式设计流程示意图（左侧输入设计目标，右侧生成方案） .....	59

## 表 目 录

表 1 科研范式的演进与特征 .....	6
表 2 科研智能主要发展阶段 .....	8
表 3 全球科研智能政策重点布局 .....	12
表 4 科研模型分类 .....	25

## 一、科研智能发展背景

本章旨在宏观研判科研智能的战略地位及历史脉络，为后续章节的深入分析奠定基础。

### （一）科研创新是经济增长的内生动力

科学研究与技术创新并非仅仅是经济活动的附属品，而是驱动经济长期增长与国家竞争力提升的核心内生动力。这一观点在现代经济学中得到了充分的理论与实证支持，其中最具代表性的是以诺贝尔经济学奖得主保罗·罗默（Paul Romer）为核心的内生增长理论。与早期新古典增长理论将技术进步视为外部给定的“黑箱”不同，内生增长理论打开了这个“黑箱”，指出经济增长源于系统内部有意识的、以市场激励为导向的投资行为，特别是对知识、人力资本和技术创新的投资。罗默的理论精髓在于，他明确了“思想”作为一种特殊经济产品的独特属性：非竞争性。一个思想或一项技术一旦被创造出来，就可以被无数人同时使用而不会被消耗，这带来了知识积累的规模报酬递增效应。例如，一家公司在研发一款新电脑时，需要花费大量人力、资金和时间，一旦研发成功，这项设计和技术就可以被反复使用，而无需重新发明。其他企业还可以在此基础上进行改进或创新，从而推动整个行业的进步。正是这种知识的非竞争性与可扩散性，使得人均产出能够实现持续增长。

内生增长理论深刻地揭示了研发（Research & Development）活动对于国家经济的重要性。宏观层面，研发投入正在成为国家竞争力的结构性变量，并呈现“高强度、头部集中、持续加码”的全球格局。

按国际可比口径最新数据统计<sup>1</sup>，2022 年排名前 8 的经济体/地区合计占全球研发总支出（Gross domestic expenditure on R&D，GERD）的 82%，其中美国约占 30%、中国约占 27%、欧盟（EU-27）约占 18%，美中合计已超过一半，反映出研发能力正在向头部经济体快速集聚并固化为战略优势。根据图 1 所示<sup>2</sup>，主要经济体研发投入强度长期上行且梯队分明，韩国、美国与日本长期保持高位，中国持续上升并加速追赶，欧盟整体相对平稳偏低，研发强度的长期分层与创新能力及产业竞争力差异呈现高度相关，并可能借助人才、平台与产业生态的累积效应，持续强化长期竞争位势<sup>3</sup>。更进一步，从政策评估视角看，国际机构也给出了可量化的增长增益：国际货币基金组织在对公共支出结构的模拟分析中指出，在发达经济体情景下，将相当于 GDP 的 1% 的公共支出从低效领域重新配置到研发，可使长期产出水平提高约 3%，这从宏观层面提供了“研发投入为何是高回报增长策略”的直观量级证据<sup>4</sup>。微观层面，企业研发投入往往能带来更高的生产率与长期竞争优势，但企业对研发的自发投入通常低于社会最优水平，其根源在于知识的可扩散性与技术外溢：企业能够内部化一部分研发收益，但研发形成的新知识会通过专利披露、工程模仿、供应链扩散、人才流动等渠道外溢，使“全社会回报”显著高于“企业账面回报”。

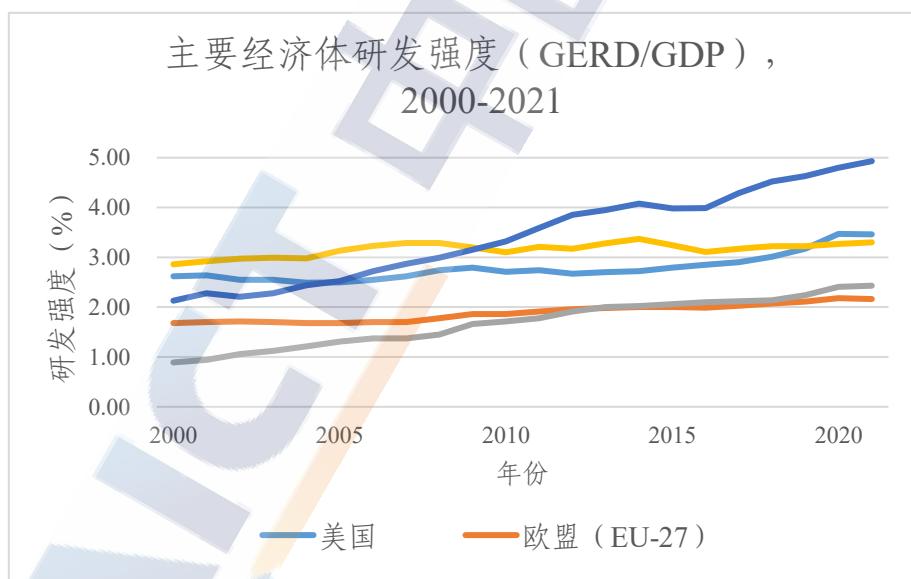
<sup>1</sup> National Science Board. Discovery: R&D Activity and Research Publications (Science and Engineering Indicators)[R/OL]. Alexandria, VA: National Science Foundation, 2025-07-23[2026-01-08]. <https://ncses.nsf.gov/pubs/nsb20257>.

<sup>2</sup> GERD 份额数据更新到 2022 年，GERD/GDP 历史数据以可获得的 2000–2021 年可比统计为基础。

<sup>3</sup> National Science Board. Research and Development: U.S. Trends and International Comparisons—Cross-National Comparisons of R&D Performance[EB/OL]. (2024-05-21) [2026-01-08]. <https://ncses.nsf.gov/pubs/nsb20246/cross-national-comparisons-of-r-d-performance>

<sup>4</sup> International Monetary Fund. Fiscal Monitor: Spending Smarter—How Efficient and Well-Allocated Public Spending Can Boost Economic Growth (October 2025)[R]. Washington, DC: IMF, 2025-10

在量级上，元分析研究显示企业研发的私有回报约在 14%，而计入外溢后的社会回报在代表性研究中可达 50%+，从而构成研发补贴、税收优惠、基础研究资助与产学研协同的坚实实证基础<sup>5</sup>。这一点也能从企业竞争格局得到更直观的佐证：欧盟委员会联合研究中心的研究显示，全球 Top 2000 研发投资企业 2024 财年合计研发投入约 1.446 万亿欧元，并覆盖全球企业资金来源研发的 90%+；按总部地区看，美国企业约占 47.1%，欧盟约占 16.2%，中国约占 16.1%，且头部企业集中度继续上升，表明研发能力正在成为企业进入并稳固全球价值链高端的关键门槛<sup>6</sup>。因此，无论从国家增长与竞争位势，还是从企业长期价值创造与外溢机制看，加大对科研创新的持续投入，已成为提升国家竞争力与推动高质量增长的核心战略选择。



来源：Discovery: R&D Activity and Research Publications

图 1 全球主要经济体研发强度示意图（2000 年至 2021 年）

<sup>5</sup> Frontier Economics. Rate of Return to Investment in R&D: A report for the Department for Science, Innovation and Technology (DSIT)[R/OL]. 2023-03[2026-01-08]. <https://www.frontier-economics.com/media/015adtpq/rate-of-return.pdf>.

<sup>6</sup> European Commission, Joint Research Centre (JRC); Directorate-General for Research and Innovation (DG RTD). The 2025 EU Industrial R&D Investment Scoreboard[R/OL]. Luxembourg: Publications Office of the European Union, 2025

## （二）科研范式的历史演进与当代局限

科研创新的方法论自身也在不断演进，已故图灵奖得主、数据库领域先驱吉姆·格雷（Jim Gray）曾提出了科学的研究的四种范式理论<sup>7</sup>，被学界与产业界广泛引用，为我们理解当前正在发生的变革提供了深刻的历史视角<sup>8</sup>。

**第一范式：经验科学（Empirical Science）**，又称实验科学。数千年前，科学源于对自然现象的直接观察、记录和描述。这一范式以实验为基础，强调经验归纳。天文学家第谷·布拉赫通过数十年如一日的观测，积累了当时最精确、最系统的行星位置数据目录，是经验科学的典范。其他典型例子还包括：卡尔·林奈建立基于形态特征的生物分类体系；法拉第的实验证实变化的磁场能产生电流，确立电磁感应现象；汤姆孙通过阴极射线实验发现电子，揭示原子内部结构。

**第二范式：理论科学（Theoretical Science）**。近几百年，随着数学工具的成熟，科学家开始构建普适性的理论模型和定律来解释和预测自然现象。约翰内斯·开普勒基于第谷的观测数据，通过数学分析发现了行星运动三定律，标志着科学从纯粹的描述走向了理论解释，这是从第一范式到第二范式的关键过渡。牛顿定律、麦克斯韦方程组以及爱因斯坦相对论是这一范式的伟大成就。

**第三范式：计算科学（Computational Science）**。20世纪下半叶，

<sup>7</sup> HEY T, TANSLEY S, TOLLE K. The fourth paradigm: data-intensive scientific discovery[M]. Redmond, WA: Microsoft Research, 2009.

<sup>8</sup> 托马斯·库恩（Thomas S. Kuhn）在《科学革命的结构》提出的科学范式是科学哲学概念，强调科学共同体的世界观会周期性更替，新范式会推翻并取代旧范式，如物理学中爱因斯坦相对论、牛顿经典力学、亚里士多德运动学逐步取代的过程。格雷的科学范式侧重科学方法论及科学实践，新方法（如数据科学）可以叠加在旧方法（如理论科学、实验科学）之上，共同存在。本报告沿用“范式”来描述科研方法论的演进。

计算机的诞生使得对复杂系统的模拟仿真成为可能。当理论方程难以解析求解时，科学家可以通过大规模数值计算来模拟现实世界中的复杂现象，例如建立全球气候模型（GCM）预测未来气候变化，使用计算流体力学（CFD）设计飞机和赛车，使用数值模拟计算核反应的临界质量。

**第四范式：数据密集型科学（Data-Intensive Science）。**进入 21 世纪，高通量测序仪、大型强子对撞机、太空望远镜等先进科学仪器以及无处不在的传感器网络，以前所未有的速度和规模产生着海量科学数据。科学研究的重心开始转向对这些庞大的数据集中探索、挖掘和发现知识。这一范式综合了理论、实验和模拟，其核心方法论涵盖了数据的采集、管理和分析的全流程。典型案例包括：斯隆数字巡天项目（SDSS）定期公开天文数据加速太空发现，人类基因组计划（HGP）完成基因测序并确立公开共享政策为全基因组研究打下基础，气候再分析数据 ERA5 长期稳定公开被广泛用于气候研究。

这四种范式并非简单的线性替代，而是一个不断累积和融合的过程。然而，随着科学研究问题的日益复杂和数据量的持续爆炸，前四种科研范式正面临着深刻的局限与挑战：**一是人类认知瓶颈。**科研文献和数据集的指数级增长速度，已经远远超出了人类研究者的阅读、处理和理解能力。据统计，仅生物医学领域的文献就以每分钟数篇的速度增长。这种“信息过载”使得研究者难以跟上领域前沿，更难以进行跨领域的知识融合与创新，导致假设生成和实验设计等关键决策环节成为科研效率的瓶颈。**二是高维复杂性。**许多前沿科学问题，如

新材料设计、药物发现和气候预测，本质上是在一个极其巨大且复杂的高维空间中进行搜索。例如，潜在的药物分子空间估计高达  $10^{60}$ ，而稳定晶体材料的可能组合更是天文数字。传统的理论指导和计算模拟方法在这种高维空间中进行“盲目”搜索，效率极低，如同大海捞针。**三是实验试错成本高昂。**传统的“假设-实验-验证”循环严重依赖物理实验。在新材料、新药研发等领域，一个完整的研究周期往往长达 10 至 20 年，耗资数亿甚至数十亿美元，且失败率极高。这种低效的试错模式已成为加速创新的主要障碍。**四是数据价值利用不充分。**21 世纪以来虽然产生了海量数据，但这些数据往往是异构、多模态且充满噪声的。如何从这些碎片化的数据中高效、自动地提炼知识、生成科学假设，并指导下一步的实验，仍然是一个巨大的挑战。数据本身并不能自动转化为知识，数据与知识之间的鸿沟亟待填补。

这些挑战共同指向一个结论：仅仅增加数据量和计算能力已不足以应对未来科研的复杂性。科研创新迫切需要引入更强大的“智能”工具，以克服人类认知的局限，驾驭高维复杂性，并实现数据、理论、计算与实验之间的高效闭环。科研智能正是在这一需求下兴起，成为推动科研范式向“第五范式”跃迁的核心引擎。

表 1 科研范式的演进与特征

范式名称	大致时期	核心研究方法	典型成果案例	核心局限
第一范式 经验科学	古代，17 世 纪前为主	观察、实验 与归纳	第谷行星观测星表， 林奈形态分类，法拉 第电磁感应，汤姆孙 发现电子	缺乏理论解 释

第二范式 理论科学	17世纪起， 20世纪中叶 为主	数学与演绎	开普勒行星三定律， 牛顿力学，麦克斯韦 方程组，爱因斯坦相 对论体系	复杂系统难 解析
第三范式 计算科学	20世纪中叶 起，下半叶 为主	计算机模拟 与仿真	全球气候模型，计算 流体力学，核反应数 值模拟	数据处理能 力不足
第四范式 数据密集 型科学	21世纪以来	大规模数据 挖掘与分析	斯隆数字巡天，人类 基因组计划，气候再 分析数据	因果性挑战 及可解释困 境

来源：中国信通院

## 二、科研智能发展历程

### （一）概念及内涵

科研智能（AI for Research and Development, AI4RD）是以人工智能（Artificial Intelligence, AI）为核心支撑，对科学研究与产业研发活动进行系统性升级的方法与工程体系<sup>9</sup>，既包括面向科研任务的模型与算法，也包括支撑其落地运行的工程平台、流程机制与治理规范。它面向“问题提出—知识获取—建模推理—计算求解—实验验证—结果沉淀与转化”的全过程，将数据与知识、模型与算法、算力资源以及实验条件纳入统一的任务组织与评估框架，形成可追溯、可复现、可持续迭代的研发流程。与把AI用于局部辅助不同，科研智能强调在科学规律与工程约束下，把“理解—生成—评估—验证”贯通起来，以更低成本、更短周期获得更高质量的知识产出与技术创新。

从工程落地看，科研智能的关键不在于叠加某一种单点能力，而

<sup>9</sup> 学术界常用科学智能（AI for Science, AI4S）指代“人工智能赋能科学发现”的研究方向：早期更多聚焦科学机器学习、代理模型等对高维学习与模拟计算的效率提升；近年随着大模型、智能体与自动化实验融合，其研究边界正扩展到覆盖“提出问题—计算求解—实验验证”的端到端流程。本文采用“科研智能”作为统称，强调以通用智能能力推动科学研究与产业研发的整体升级。但在引用部分政策文件、学术研究等内容时，仍会按照其原始出处使用科学智能/AI4S等称呼，在此特别澄清。

在于构建可复用、可扩展的“科研流水线”。一方面，通过数据治理、知识表达与标准化接口，把分散的文献、实验与仿真信息转化为可调用的资产；另一方面，通过工作流编排、资源调度与评测反馈，把计算模拟、实验执行与结果评价联动起来，使研究过程能够被量化评估、自动化执行并在反馈中持续优化。由此，科研活动可由依赖个人经验串联的线性流程，转向可并行组织、可规模化迭代的系统化流程。

在能力层级上，科研智能推动科研能力由“加速计算”迈向“生成设计”直至“部分自主发现”。其一，面向仿真与计算密集任务，AI 可学习历史模拟数据构建代理模型，在一定精度约束下显著降低计算成本，支撑更大规模的参数探索与优化。其二，面向研发设计问题，生成式模型可在目标约束下提出候选结构或方案，促进从“先结构后验证”的正向流程向“按指标反推方案”的逆向探索转变。其三，面向实验环节，通过将 AI 决策与自动化实验平台、仪器系统以及安全与质量控制机制集成，可形成“设计—执行—分析—再设计”的闭环实验流程，推动从人机协同向更高程度的自动化探索演进。

## （二）发展历程

科研智能的发展并非一蹴而就，而是与 AI 技术自身的浪潮紧密相连。回溯其发展，可概括为“三步走”演进路径：从“辅助分析”，到“深度突破”，再到“生成与验证闭环”（见表 2）。本报告将其发展历程划分为以下三个主要阶段。

表 2 科研智能主要发展阶段

阶段	关键特征	代表性案例	阶段性价值
----	------	-------	-------

<b>阶段一： 辅助分析 (1960 年代-2011 年)</b>	规则推理期：规则推理/专家系统 数据挖掘期：传统机器学习做数据挖掘	DENDRAL (分子结构推断) MYCIN (医学诊断) 人类基因组计划 (生物信息学) SDSS (天文星系分类)	推动科研从“经验驱动”迈向“数据与知识驱动”，形成早期可复用的方法与工具链
<b>阶段二： 深度突破 (2012 年-2022 年)</b>	深度学习驱动；两条路线并行：机理-数据融合 (灰盒) 与纯数据驱动 (黑盒)	PINNs (物理信息神经网络) AlphaFold 2 (蛋白质结构预测)	深度学习开始攻克关键科学问题，实现高精度建模与预测，形成里程碑式突破
<b>阶段三： 生成验证 (2023 年-至今)</b>	生成式模型/科研大模型+自动化实验室闭环	RFdiffusion (蛋白质逆向设计) A-Lab (自主发现合成化合物) MatterGen (材料逆向设计)	从“预测/求解”迈向“生成假设与候选+验证迭代”，推动科研过程闭环化与自动化

来源：中国信通院

**阶段一：辅助分析阶段（1960 年代-2011 年）。**在长达半个世纪的早期探索中，AI 始终扮演着科学家的“辅助工具”而非合作伙伴。这一阶段本身也呈现出两个清晰的子阶段：一是规则推理期（1960 年代-1980 年代），科学家试图将人类的显性知识（如化学规则）编码为知识库，让机器模拟专家的逻辑推理。其标志性成果是 1965 年的 DENDRAL 项目<sup>10</sup>，它利用化学知识库来推断分子结构，被视为首个 AI 成功赋能科研的应用。此后出现的 MYCIN 系统<sup>11</sup>在医学诊断中应用规则库进行血液感染识别与抗生素推荐，展示了人工智能在科研决策中的潜力，同时暴露出知识获取困难、推理不确定等问题。二是数

<sup>10</sup> Lindsay R K, Buchanan B G, Feigenbaum E A, et al. DENDRAL: a case study of the first expert system for scientific hypothesis formation[J]. Artificial intelligence, 1993, 61(2): 209-261.

<sup>11</sup> BUCHANAN B G, SHORTLIFFE E H. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project[M]. Reading, MA: Addison-Wesley, 1984.

**据挖掘期（1990 年代-2011 年），90 年代后，高通量实验平台和开放科学数据库（如 GenBank）的出现，产生了传统统计学难以处理的海量数据。此时，传统机器学习（如支持向量机、贝叶斯网络）开始被大规模用于数据挖掘，成为科学家的“数据分析助手”。人类基因组计划首次系统地揭示了人类基因组的全貌，催生了“生物信息学”这一新兴交叉学科，并推动传统机器学习算法在基因识别、序列比对与功能预测中的广泛应用。与此同时，天文学领域的斯隆数字巡天计划（Sloan Digital Sky Survey）自 2000 年起持续发布大规模观测数据，推动了使用机器学习自动分类海量星系图像。**

**阶段二：深度突破阶段（2012 年-2022 年）。**2012 年深度学习在图像识别领域的爆发也成为科研智能发展的重要转折点，得益于强大的非线性拟合能力和大规模图形处理器（Graphics Processing Unit, GPU）并行计算，AI 的角色从“分析数据”转变为“解决核心问题”。在这一阶段，AI 赋能的科研探索呈现出“黑盒”与“灰盒”并行的两条路径。“灰盒”路径强调机理-数据融合，以科学机器学习（SciML）为代表，标志性成果是 2017 年提出的物理信息神经网络（PINNs）<sup>12</sup>，它巧妙地将物理定律（即偏微分方程）作为损失函数嵌入神经网络，实现了在“小数据”情况下对复杂物理系统的精准建模。2019 年，美国能源部高性能计算与先进科学研究中心（ASCR）发布《科学机器学习核心技术报告》<sup>13</sup>，正式将 SciML 确立为基础研究优先方向，

<sup>12</sup> RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. Journal of Computational Physics, 2019, 378: 686-707.

<sup>13</sup> U.S. DEPARTMENT OF ENERGY, OFFICE OF SCIENCE, ASCR. Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence[R/OL]. 2019-02[20

标志着方法学层面的制度化起点。“黑盒”路径聚焦纯数据驱动，这一路径的巅峰成就，也是整个阶段的里程碑，是 2020 年的 AlphaFold 2<sup>14</sup>。DeepMind 团队利用图网络和注意力机制，将蛋白质结构预测精度提升至接近实验分辨率水平，解决了困扰生物学界 50 年的蛋白质折叠问题，它表明深度学习有能力攻克科学中最基本、最困难的问题之一。

**阶段三：生成验证阶段（2023 年至今）。**2023 年以来，在生成式 AI、科研大模型和自动化实验室的共同推动下，AI 不再局限于分析已有数据或解决已知问题，而是开始主动生成全新的、有价值的科学假设、分子结构和材料，推动科研流程逐步形成“生成—验证—迭代”的闭环。这一阶段呈现出两条相互强化的路径：一是生成模型驱动的候选生成，面向特定约束直接产出可筛选的分子与材料候选。生命科学领域，RFdiffusion（2023 年）通过扩散式生成模型在蛋白骨架空间进行去噪采样，可按拓扑或结合位点等约束生成全新蛋白，并以实验表征验证其结构与功能，体现按目标生成的能力<sup>15</sup>。材料领域，MatterGen（2025 年）将扩散模型扩展到周期晶体，联合生成原子类型、坐标与晶格参数，并可面向逆向设计任务按性质约束微调，直接产出稳定且多样的无机材料候选<sup>16</sup>。二是自动化实验室驱动的闭环验证，把“文献与数据库 + 模型生成/筛选 + 主动学习 + 机器人合成表

26-01-07]. DOI:10.2172/1478744.

<sup>14</sup> JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold [J]. Nature, 2021, 596: 583-589. DOI:10.1038/s41586-021-03819-2.

<sup>15</sup> Watson J L, Juergens D, Bennett N R, et al. De novo design of protein structure and function with R Fdiffusion[J]. Nature, 2023, 620(7976): 1089-1100.

<sup>16</sup> ZENI C, PINSLER R, ZÜGNER D, et al. A generative model for inorganic materials design[J]. Nature, 2025, 639(8055): 624-632.

征”贯通为连续迭代流程。加州大学伯克利分校的自动化实验室 A-Lab（2023 年）在约 17 天连续运行中从 58 个目标中实现 41 种化合物，验证了“生成—验证—迭代”闭环在材料发现中的可行路径<sup>17</sup>。

### 三、科研智能政策举措

科研智能被普遍视为影响未来科技竞争力的重要方向，已引起全球主要经济体的高度重视。各国政府正通过制定国家战略、启动重大计划、增加研发投入等方式，系统性布局这一新兴赛道。尽管多数国家的综合性 AI 战略都包括了对科研的支持，但专门针对科研智能的体系化的政策部署正成为新的趋势<sup>18</sup>。

表 3 全球科研智能政策重点布局

经济体	顶层设计型政策	资源导向型政策	专项及组织型政策
美国	“创世纪计划”(2025): 由能源部牵头，整合国家资源，建立统一 AI 科学实验平台	NAIRR 计划(2024): NSF 牵头，构建国家 AI 研究资源，推动 AI 研发“民主化”	国家 AI 研究院计划：NSF 旗舰投资，已资助多个交叉领域研究所
欧盟	《科学人工智能战略》(2025): 以欧洲科学人工智能资源为抓手，统一调配全欧资源	欧洲开放科学云(EOSC): 整合成员国数据与平台，构建跨学科科学数据空间	地平线欧洲、欧洲研究委员会、欧洲创新委员会及欧洲伙伴关系已形成系统化资金支持体系
中国	“人工智能+”行动(2025): 将“AI+科学技术”列为首要行动，加速发现与范式创新	国家超算互联网：将全国各地超算中心连接成一体化的算力网络	科技部及自然科学基金委以“项目群+平台化”组织化投入

<sup>17</sup> Szymanski N J, Rendy B, Fei Y, et al. An autonomous laboratory for the accelerated synthesis of novel materials[J]. Nature, 2023, 624(7990): 86-91.

<sup>18</sup> 董昊, 魏凯. 科研智能：政策、技术及应用[J/OL]. 大数据, 1-15[2026-01-09]. <https://link.cnki.net/urlid/10.1321.G2.20250709.1501.010>.

其他	<p>1.英国：发布《AI 科学战略》，旨在利用 AI 从根本上改变科学发现的本质</p> <p>2.韩国：《AI+科学技术促进方案》(2025)，推动研发范式转型</p>	<p>韩国：启动第六代国家超算建设，构建国家科研数据平台</p>	<p>1.日本：JST 资助材料与生命科学 AI 融合研究</p> <p>2.韩国：启动新药研发、材料科学等长期项目</p>
----	--	----------------------------------	--

来源：中国信通院

## （一）美国

美国作为科技创新大国，以国家竞争力与科技主导权为牵引，把人工智能从“科研工具”升级为“科研新底座”，通过政府主导的系统工程重塑科学发现与工程研发的全链条。

**一是国家顶层动员与国家工程化牵引。**白宫于 2025 年 11 月发布行政令，启动“创世纪计划”（Genesis Mission），将 AI 赋能科研上升为国家级工程。该计划由能源部牵头实施，整合联邦层面的算力与数据资源，建设统一的 AI 科学实验平台，面向生物技术、关键材料等重点领域训练科学基础模型，并与自动化实验室衔接形成闭环，力求在十年内显著提升联邦科学与工程研发效率<sup>19</sup>。其政策信号在于从“规则与治理导向”转向“算力—数据—重大任务场景”一体化动员，并以“曼哈顿计划式”国家工程叙事强化战略紧迫性与资源聚焦。

**二是全国性共享科研基础设施的普惠供给。**美国国家科学基金会（NSF）于 2024 年初牵头启动国家人工智能研究资源（NAIRR）计划试点，目标是汇聚并开放共享算力、数据、软件工具与模型资源，

<sup>19</sup> The White House. Executive order: launching the Genesis Mission[EB/OL]. (2025-11-24)[2026-01-06]. <https://www.whitehouse.gov/presidential-actions/2025/11/launching-the-genesis-mission/>.

降低科研团队使用 AI 的门槛，推动 AI 研发的“民主化”，根据公开信息统计，截至 2025 年 11 月 NAIRR 已支持 540 余个项目，其中约 25% 的项目聚焦科研智能。

**三是以长期研究网络构建跨机构的组织机制与生态。**NSF 牵头的国家 AI 研究院是美国在 AI 赋能科研和产业应用领域的旗舰级长期投入载体，其特点是通过多年度、跨机构的研究院网络，把 AI 方法、数据资源与学科问题长期耦合，形成稳定的人才培养、工具平台沉淀与跨学科协同机制。截至 2025 年底，该计划已通过五轮主要批次（2020 年、2021 年、2023 年、2024 年及 2025 年），累计资助了约 30 所国家 AI 研究所，包括 AI 赋能物理、天文学、气候科学、材料、合成生物学等领域。

## （二）欧盟

欧盟以“欧洲一体化协同”为主线，致力于把分散在成员国与机构间的算力、数据、人才与资金编织成可共享、可复用的科研能力网络，推动科研范式从单点突破走向跨域协同创新。

**一是确立战略牵引与统筹机制。**欧盟于 2025 年 10 月发布“科学人工智能战略”（A European Strategy for Artificial Intelligence in Science），提出以人工智能系统性加速科学发现与技术创新，并以“欧洲科学人工智能资源（Resource for AI Science in Europe，RAISE）”作为核心抓手，统筹整合欧洲范围内的算力、数据、人才与科研资金，形成面向科研群体的共享能力供给。战略明确以试点方式推进 RAISE 建设，计划投入约 1.07 亿欧元启动相关工作，并通过后续更大规模

投入持续强化科研算力获取与高质量科学数据建设<sup>20</sup>。

**二是建设关键资源底座。**在算力侧，欧盟依托欧洲高性能计算联合体（EuroHPC）推进“人工智能工厂（AI Factories）”与“人工智能超级工厂（AI Gigafactories）”等基础设施布局，提升面向科研的大模型训练、科学计算与跨学科任务的算力供给能力，并探索更适配科研项目的访问与优先机制。在数据侧，欧盟以欧洲开放科学云（European Open Science Cloud，EOSC）为牵引建设开放、可信、跨学科的科学数据空间；2024年10月上线的EOSC欧盟核心节点（EOSC EU Node）作为联邦化开放科学云的统一入口与参考节点，推动各成员国与学科数据/服务节点互联互通，为科研智能提供可发现、可访问、可复用的数据与服务底座。

**三是提供资金与创新工具箱。**欧盟已形成较为系统的资金支持体系，地平线欧洲（Horizon Europe）作为总盘资金广泛支持人工智能基础研究与应用研究；欧洲研究委员会（European Research Council，ERC）聚焦研究者驱动的前沿探索；欧洲创新委员会（European Innovation Council，EIC）面向具备商业潜力的突破性技术创新与初创/中小企业规模化；欧洲伙伴关系（European Partnerships）则通过公私协同组织联合体攻关，面向特定技术方向与行业挑战形成长期合作与资源汇聚。这套“科研—转化—产业化”的资金支持链条为科研智能的持续投入与扩散提供制度性保障。

<sup>20</sup> European Commission. AI in Science Strategy: Paving the way for European AI for Science Resources [R/OL]. (2025-10-08)[2026-01-06]. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52025DC0724>.

### （三）中国

中国同样高度重视科研智能的发展，以国家战略目标为牵引，把人工智能嵌入科学与研发全流程，通过“算力与数据底座—组织化项目群—区域试点示范”的联动机制，推动科研范式与创新链条同步升级。

**一是国家战略牵引**，以“人工智能+科学技术”作为国家行动的优先方向，强调科研范式与创新链条的系统重构。2025年8月，国务院印发《关于深入实施“人工智能+”行动的意见》，将“人工智能+科学技术”列为六大重点行动之首，明确指出：一是加快探索人工智能驱动的新型科研范式，加速“从0到1”重大科学发现进程；二是推动人工智能驱动的技术研发、工程实现、产品落地一体化协同发展，加速“从1到N”技术落地和成果转化；三是推动哲学社会科学研究方法向人机协同模式转变，拓展对人工智能影响的研究与治理能力建设<sup>21</sup>。

**二是关键资源底座**，突出“科研友好型算力供给能力”，同时补齐科学数据供给短板。在算力侧，2024年科技部推动“国家超算互联网”建设，旨在以算力网络连接全国超算中心并形成一体化算力服务平台，融合超算与智算算力，强化跨区域调度与服务能力，降低科研与产业使用门槛。同时，科技部推进“国家新一代人工智能公共算力开放创新平台”布局，已出现“9家获批建设、16家获批筹建”的梯队式供给格局，面向科研与产业开放普惠算力服务。在数据侧，2019

<sup>21</sup> 国务院. 国务院关于深入实施“人工智能+”行动的意见[EB/OL]. (2025-08-27)[2026-01-07]. [https://www.gov.cn/zhuengce/content/202508/content\\_7037861.htm](https://www.gov.cn/zhuengce/content/202508/content_7037861.htm)

年，我国正式建成 20 个国家科学数据中心和 30 个国家生物种质与实验材料资源库，但总体上仍面临跨部门、跨区域、跨学科的统一汇聚与共享服务能力不足的问题，需要与算力网络协同推进“可发现、可获取、可复用”的高质量科学数据集建设与开放机制。

**三是专项与组织机制**，以“项目群+平台化”组织化投入，推动从方法突破到工程落地的持续迭代。科技部层面，通过“人工智能驱动的科学研究”专项部署等机制，推动面向重大科学问题的模型与算法创新，并强调平台化与工程化落地。国家自然科学基金委员会层面，形成“方法论供给+任务牵引”的项目组合：既通过重大研究计划聚焦可解释、可通用等下一代人工智能方法，夯实跨学科通用能力，也通过专项项目群面向工程科学前沿探索、肿瘤精准“智疗”、材料科学模型驱动发现、复杂系统智能表征与建模等方向组织攻关。

**四是地方协同**，以行动计划与重点赛道为抓手，承担“场景组织+要素汇聚+试点示范”的加速器角色。AI+科研方面，北京、浙江等地推出“人工智能+科学”行动计划类政策文件，以区域算力底座、数据底座、模型与平台能力、人才与生态体系为主线，组织高校院所、头部企业与平台机构协同建设，强化高价值科研场景与试点任务牵引。在 AI+材料方面，北京、上海等地围绕“人工智能+材料/材料智能引擎”等发布专项政策文件，突出产业链与科研链协同：以材料研发流程为主线，推动“计算/模拟—数据—模型—实验”贯通，布局高通量与智能化实验平台、材料领域垂类模型与工具链，形成可持续迭代的工程化能力，并通过示范应用扩散到新材料重点方向。

## （四）其他国家

其他主要国家也在积极加码科研智能政策，纷纷将“AI 赋能科研”视为未来科技竞争的核心方向，围绕顶层政策设计、算力基础设施、科研数据开放和跨学科 AI 应用等方面推出系统性举措，形成多层次、协同化的国际发展格局。

英国科学、创新与技术部于 2025 年 11 月发布《人工智能促进科学发展战略》（AI for Science Strategy），将科研智能定位为提升国家科研竞争力与产业增长的关键抓手。战略提出两项目标：一是发展人工智能驱动科学的研究的前沿能力，二是确保英国继续保持全球科学领导力；并以“数据—算力—人才与文化”三大支柱推进落地。政策聚焦五大优先方向：工程生物学、聚变能源、材料科学、医学研究与量子技术，强调以国家级资源杠杆加速科研范式变革。战略引入任务牵引机制，已公布首个任务——到 2030 年实现“100 天研发可进入临床试验的药物”<sup>22</sup>。

日本则通过国家战略层面明确提出科研智能的发展方向。文部科学省在《2024 科学技术与创新白皮书》及配套政策文件中首次将“科研智能”作为重点任务，提出要构建融合算力、网络与数据的一体化科研基础设施体系，包括国家超级计算机系统、科研信息网络及科研数据基础设施<sup>23</sup>。同时，该部通过科学技术振兴机构（JST）资助多项跨学科研究计划，如“面向材料与生命科学的人工智能融合研究”等，

<sup>22</sup> Department for Science, Innovation and Technology (DSIT). AI for Science Strategy[R/OL]. (2025-11-20)[2026-01-06]. <https://www.gov.uk/government/publications/ai-for-science-strategy>.

<sup>23</sup> 文部科学省. 令和 6 年版科学技術・イノベーション白皮書: AI が変える科学技術・イノベーション[R/OL]. (2024-06-11)[2026-01-06]. [https://www.mext.go.jp/b\\_menu/hakusho/html/hpab202401/index.html](https://www.mext.go.jp/b_menu/hakusho/html/hpab202401/index.html).

依托理化研究所、北海道大学等核心机构，在化学、材料、生物与环境等领域推动 AI 应用深化。2024 年 4 月，美日首脑会谈还将“AI for Science”合作写入联合声明，标志科研智能上升为国际合作新焦点。

韩国在 AI 赋能科研领域同样大幅提速。2025 年 3 月，韩国以跨部门形式公开《AI+科学技术促进方案》，明确提出以 AI 驱动科研体系转型，顺应全球“研发范式向 AI 中心转变”的趋势，加快 AI 在科研活动中的系统嵌入<sup>24</sup>。作为支撑，韩国科学技术信息研究院（KISTI）于 2025 年 5 月启动第六代国家超级计算机建设，面向“AI+科学技术”场景提供海量计算能力。KISTI 还负责运营国家科研高速网络，并推动建设国家科研数据平台，形成服务科研智能的基础设施底座。在应用层面，韩国围绕新药研发、材料科学等领域启动大型长期项目，推动 AI 科研平台化与持续化发展，成为本轮人工智能发展的重要策源地。

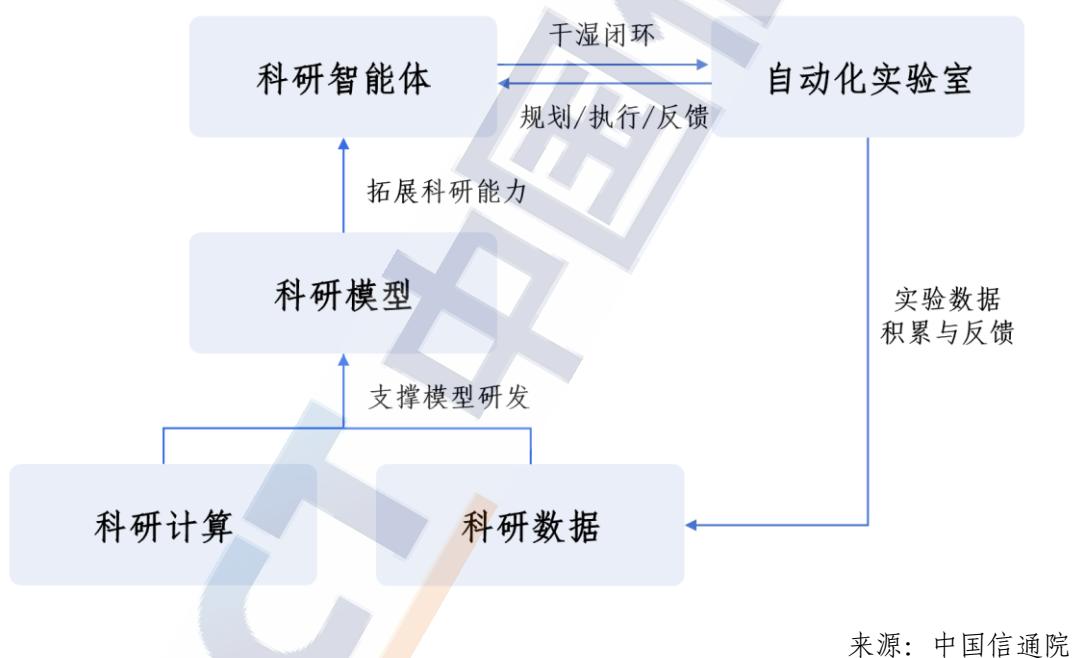
加拿大则从科研算力供给角度强化布局。2025 年，加拿大数字研究联盟（Digital Research Alliance of Canada）启动“国家 AI 计算——快速部署计划”（National AI Compute – Rapid Deployment），目标是为科研人员提供可快速接入的 AI 计算资源，以满足科研界对 AI 算力爆发式增长的需求。这一计划标志着加拿大正式将 AI 计算纳入国家科研基础设施体系，为科研智能提供坚实的算力支撑<sup>25</sup>。

<sup>24</sup> 韩国政府相关部委联合. 为实现全球科技强国的 AI+S&T 振兴方案 [R/OL]. (2025-03-12)[2026-01-06].

<sup>25</sup> Digital Research Alliance of Canada. The Alliance invites proposals to expand artificial intelligence compute resources: National AI Compute – Rapid Deployment initiative[EB/OL]. (2025-07-02)[2026-01-06]. <https://www.alliancecan.ca/latest/news/alliance-invites-proposals-expand-artificial-intelligence-compute-resources>.

## 四、科研智能关键技术

本报告基于基础通用性原则，聚焦五大技术领域展开讨论，即科研数据、科研计算、科研模型、科研智能体和自动化实验室。其中，科研数据与科研计算共同支撑科研模型的研发；科研智能体在模型基础上进一步拓展科研能力，并通过自动化实验室实现“干湿闭环”，打通数字世界和物理世界。上述技术正在快速迭代创新，衍生出一系列科研工具及服务，补充甚至替代部分传统科研工具，推动科研工作高效开展<sup>26</sup>。



来源：中国信通院

图 2 科研智能关键技术示意图

### （一）科研数据

科研数据指在科研过程中产生、收集和应用的各类数据资源，以及围绕数据获取、处理、管理和应用形成的一系列技术与规范体系。其涵盖实验观测数据（如传感器读数、测量记录）、模拟计算数据（如

<sup>26</sup> 董昊, 周景才. 科研智能关键技术研究[J]. 信息通信技术与政策, 2025, 51(8): 42.

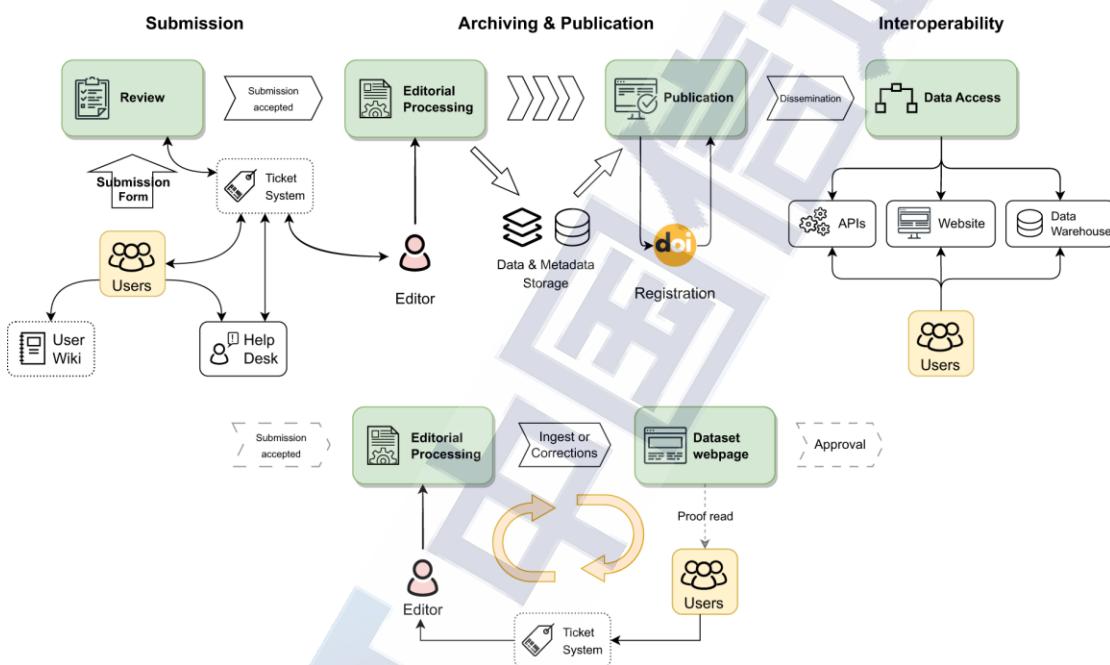
数值仿真输出）、文献与知识库数据（如论文、专利等文本）以及衍生的数据库和知识图谱等。在大模型时代，科研数据已成为关键要素和生产资料，核心要求也从“可存可取”进一步走向“可发现、可获取、可互操作、可复用（Findability, Accessibility, Interoperability, and Reusability, FAIR）<sup>27</sup>”，并强调对算法、工具与工作流等数据产生过程的可追溯与可复现支持。科研数据主要涉及数据采集与生成、科研数据治理、科研知识图谱等关键技术。

**其一，数据采集与生成。**该过程通过多渠道获取原始科研数据并产生新数据，包括实验观测数据采集、模拟仿真数据产生等。目前，业界借助智能采集软件、自动化实验室等技术实现数据采集的自动化和高通量。例如，美国劳伦斯伯克利国家实验室（简称“伯克利实验室”）开发的 gpCAM 软件，利用贝叶斯等算法优化采样准则，在模拟或实验过程中实时更新不确定度并确定下一测点，可将二维材料量子特性显微成像的时间从 23 天缩短至 8 小时，已被广泛应用于模拟与实验数据的自主获取。同时，在材料、化学与工程仿真场景中，高通量科学计算（如基于密度泛函的第一性原理计算、分子动力学模拟等）已成为重要数据来源：通过标准化工作流批量生成高精度模拟数据，并与实验数据互校、互补，可为模型训练与评测提供可控的标注数据，也为“仿真—学习—实验”闭环优化提供数据底座。

**其二，科研数据治理。**这是围绕科研数据全生命周期所进行的管理、组织和规范化工作，旨在保证数据的质量、安全、合规性与高效

<sup>27</sup> Wilkinson M D, Dumontier M, Aalbersberg I J J, et al. The FAIR Guiding Principles for scientific data management and stewardship[J]. Scientific data, 2016, 3(1): 1-9.

利用，主要涵盖数据分类分级、数据清洗与整合、元数据与语义管理、权限与安全控制以及标准制定等技术手段。完善的数据治理架构可以显著提升科研数据的利用价值。以地球系统科学领域为例，数据期刊与权威数据仓库协同的数据发表机制被广泛采用：研究者将数据提交至 PANGAEA 等数据仓库并按要求完善元数据与规范流程，形成可长期保存与可引用的数据资产<sup>28</sup>。



来源：PANGAEA - Data Publisher for Earth & Environmental Science

图 3 PANGAEA 从数据提交到发布使用的工作流程示意图

**其三，科研知识图谱。**该技术将科研中的各种知识要素（实体）及其关系以图结构组织与表示，支撑知识检索、关联发现与推理分析。构建知识图谱通常涉及实体抽取与消歧、关系抽取、知识融合、语义对齐与推理等技术，知识更新和表示学习也是重要技术分支。随着大模型在信息抽取、语义理解与跨模态对齐方面能力增强，知识图谱正

<sup>28</sup> Felden J, Möller L, Schindler U, et al. PANGAEA-data publisher for earth & environmental science[J]. Scientific Data, 2023, 10(1): 347.

在从“人工规则驱动”走向“数据—模型协同构建”，并成为连接文献、数据集、代码与实验记录的语义枢纽。业界已涌现多个相关项目。

例如美国艾伦人工智能研究所建立的 Semantic Scholar 平台提供 AI 驱动的学术知识图谱服务，目前涵盖 2.14 亿篇论文、24.9 亿条引用和 7900 万份作者信息，为文献推荐、趋势分析与知识发现提供支撑<sup>29</sup>。

## （二）科研计算

科研计算支撑科研领域活动的高性能计算资源和技术体系。相较于通用信息技术算力，科研计算不仅涵盖传统科学计算所需的高精度数值模拟算力，还包括训练科研大模型所需的 AI 算力。硬件层面包括由中央处理器（Central Processing Unit, CPU）、GPU、专用集成电路（Application-Specific Integrated Circuit, ASIC）等计算芯片、网络存储、高速网络以及配套设施组成的计算集群；软件层面包括高性能计算操作系统、资源调度系统、并行计算框架以及科学计算与 AI 计算软件库等。总体而言，科研计算主要包括并行效率优化、异构计算与资源调度等关键技术。其中，并行效率关注算法并行分解、通信开销与负载均衡，以提升强/弱扩展能力；异构计算关注多类型算力单元协同以提升性能与能效；资源调度则通过高效管理与分配计算资源以支撑多用户、多任务的稳定运行。

异构计算已成为 AI 训练、推理与部分计算密集型仿真场景中的主导形态，其中最典型的是 CPU+GPU 协同架构：CPU 负责通用控制

<sup>29</sup> Semantic Scholar. Semantic Scholar Academic Graph API[EB/OL]. [2026-01-05]. <https://www.semanticscholar.org/product/api>

与串行/中等并行任务，GPU 承担大规模并行与高吞吐计算，两者协同加速任务执行效率。其核心挑战不仅在于“采用何种硬件、如何实现并行”，更在于任务映射与数据移动以及编程模型与软件生态适配。在编程生态方面，英伟达并行计算框架 CUDA 在深度学习与大量高性能计算（High Performance Computing, HPC）加速场景中仍具有显著优势，同时，业界也在推动面向多硬件的编程与运行时体系建设，以降低对单一生态的锁定风险。同时，CPU 与 GPU 正通过更紧耦合的封装与互连走向“统一系统级算力”，以减少应用迁移与运行过程中的内存搬迁开销并提升带宽与能效，例如 AMD Instinct MI300A 采用芯粒（Chiplet）与 3D 堆叠将 CPU 芯粒、GPU 芯粒与高带宽存储器（HBM）集成于单一封装，英伟达 GH200 Grace Hopper 将 Grace CPU 与 H100 GPU 在单一模块中通过高速互连实现更紧密协同，从而同时服务 HPC 与 AI 负载。

资源调度面向多用户、多任务环境，目标是在满足作业需求的同时提升资源利用率、缩短完成时间并维持公平性，重点解决“在哪运行、何时运行、运行哪些任务”的问题。工程上主要形成批处理调度（典型用于 HPC 作业队列管理，如 Slurm）与云原生调度（典型用于服务型负载管理，如 Kubernetes 及其面向批任务/AI 的扩展）两条路径，并在异构算力需求快速增长的背景下呈现融合演进的探索，通过统一资源抽象、混合调度策略与弹性伸缩等机制，实现对 CPU/GPU 等异构资源的更高效管理与优化。

### （三）科研模型

科研模型是指正从通用大语言模型向“面向科研知识表达、科研推理与科研对象表征”的专用/增强模型体系演进，科研模型的统一标准分类尚未完全固化，考虑到“科研数据资产的主要载体与表征形态”在综述性论文中反复出现，也更便于工程落地，本报告参考该维度将科研大模型划分为三类，科研大语言模型侧重文档与记录的理解、归纳与论证，领域科研大模型侧重科学对象结构化表征的预测、生成与优化，多模态科研大模型侧重跨模态对齐与联动，打通文本证据与结构对象。

表 4 科研模型分类

科研模型类别	核心对象/目标	关键技术要点	典型应用及代表性工作
科研大语言模型 (文档与记录类)	面向论文/专利/标准、实验记录等“文档+记录”，强化长文档理解归纳、证据链推理、结构化表达	高质量科学语料继续预训练+科研任务指令微调/对齐	文献综述、条款归纳、记录整理、科研写作/问答；SciGLM（2024）、SciLitLLM（2025）
领域科研大模型 (结构化对象表征类)	面向序列/图/连续场等科学对象表征，实现预测+生成+优化	核心在“表示—约束—可验证”：语法/结构有效+满足性质/功能等目标约束的可控建模，并可通过仿真评估或实验进行外部核验	材料/化学/生物分子/天气等：性质预测、候选生成与优化；AlphaFold 3（2024）、ESM3（2025）、GraphCast（2023）、NeuralGCM（2024）、GenCast（2025）、GNOME（2023）
多模态科研大模型	对齐文本与结构/谱图/图像/场	常用“模态编码器+基础模型”对齐	文献+结构/图像联合检索、结构解释问答、跨

(跨模态对齐联动类)	等，实现跨模态检索、理解、生成/编辑，打通“文本证据—结构对象”	架构；通过投影器/查询模块与参数高效适配降低迁移成本	模态编辑；MolCA（2023）、MolLM（2024）
------------	----------------------------------	----------------------------	------------------------------

来源：中国信通院

**科研大语言模型**是指以论文、专利、标准等科学文献，以及实验记录、研发日志、数据表单等“文档+记录”类资产为主要处理对象，核心目标是提升科学概念理解、长文档归纳、证据链推理与面向科研任务的结构化表达能力。其典型技术路径仍是“高质量科学语料继续预训练+科研任务指令微调/对齐”，但相较于通用对话模型更强调两点：一是对文献与记录数据进行结构化解析与清洗，二是在长上下文条件下维持结论一致性、边界意识与不确定性表达。该类模型适用于文献调研与综述、专利与标准条款归纳、实验/研发记录结构化整理、科研写作与解释型问答等知识密集型工作。代表性工作如 SciGLM（2024年）通过自反思指令标注与微调提升科学/数学任务能力<sup>30</sup>；SciLitLLM（2025年）围绕科学文献理解构建系统化适配策略与实证评测，体现文档与记录类科研模型从“能对话”向“能读懂科研文献并形成可复用结论”演进<sup>31</sup>。

**领域科研大模型**以“科学对象的结构化表征数据”为核心建模对象，覆盖三类典型形态：一是离散符号序列（如分子反应表达、蛋白/核酸序列、科学计算代码等），二是图与网络（如分子2D图、知识

<sup>30</sup> Zhang D, Hu Z, Zhoubian S, et al. Sciglm: Training scientific language models with self-reflective instruction annotation and tuning[J]. arXiv preprint arXiv:2401.07950, 2024, 4.

<sup>31</sup> LI Sihang, HUANG Jin, ZHUANG Jiaxi, et al. SciLitLLM: How to Adapt LLMs for Scientific Literature Understanding[C]//Proceedings of the International Conference on Learning Representations (ICLR 2025). 2025.

图谱、相互作用网络等），三是连续信号与场（如谱图、显微/遥感图像、三维几何与时空场、仿真网格输出等）。该类模型的关键在于“表示—约束—可验证”：既要保证生成或预测结果在语法/结构上有效，又要能在功能、性质、活性等目标约束下实现可控建模与优化，并能够通过数据库对照、仿真评估或后续实验进行外部核验。其适用场景覆盖化学、材料、生物分子、地球科学与工业仿真等方向的性质预测、结构理解、候选生成与优化、跨尺度建模等任务。近年来取得多项突破性进展，在生物分子方向，AlphaFold 3（2024 年）将结构预测扩展到蛋白—核酸一小分子等更一般的生物分子相互作用体系，显著提升复合体层面的结构建模能力<sup>32</sup>；ESM3（2025 年）以“序列—结构—功能”联合建模与生成推动蛋白基础模型从表征走向设计<sup>33</sup>。在连续场方向，GraphCast（2023 年）代表数据驱动中期天气预报的里程碑<sup>34</sup>，NeuralGCM（2024 年）体现“可微动力学+机器学习组件”的混合建模路径<sup>35</sup>，GenCast（2025 年）则强调概率集合预报与不确定性表达，推动模型从“单次预测”走向“可用的集合预报”<sup>36</sup>。在材料方向，GNOME（2023 年）以图网络与高通量筛选为核心，在新材料发现规模上产生显著影响，但其整体实践往往包含“模型+搜索/验证流程”，严格意义上并非纯模型<sup>37</sup>。

<sup>32</sup> Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3[J]. Nature, 2024, 630(8016): 493-500.

<sup>33</sup> Hayes T, Rao R, Akin H, et al. Simulating 500 million years of evolution with a language model[J]. Science, 2025, 387(6736): 850-858.

<sup>34</sup> Lam R, Sanchez-Gonzalez A, Willson M, et al. Learning skillful medium-range global weather forecasting[J]. Science, 2023, 382(6677): 1416-1421.

<sup>35</sup> Kochkov D, Yuval J, Langmore I, et al. Neural general circulation models for weather and climate[J]. Nature, 2024, 632(8027): 1060-1066.

<sup>36</sup> Price I, Sanchez-Gonzalez A, Alet F, et al. Probabilistic weather forecasting with machine learning[J]. Nature, 2025, 637(8044): 84-90.

<sup>37</sup> Merchant A, Batzner S, Schoenholz S S, et al. Scaling deep learning for materials discovery[J]. Natur

**多模态科研大模型**面向科学信息多模态共存的现实需求，对齐并联合建模至少两类“原子模态”（如文本与分子图/三维结构、文本与谱图/图像、文本与时空场等），实现跨模态理解、检索、生成与编辑。其关键技术通常采用“模态编码器+基础模型”的对齐架构，并通过投影器/查询模块与参数高效适配方法（如低秩适配 LoRA）降低跨模态迁移成本；工程难点主要在高质量跨模态配对数据构建（例如结构一文本、图像一描述、谱图一条件/结论）以及在对齐过程中同时保持科学约束与可用的生成/推理能力。该类模型适用于“文献+结构/图像”联合检索、结构解释型问答、跨模态编辑生成，以及连接文本证据与结构对象的研发协同场景，是提升“知识—对象”联动效率的重要路径。代表性工作如 MolCA（2023 年）系统评测分子图—语言对齐在分子描述生成、命名与检索等任务上的效果<sup>38</sup>；MolLM（2024 年）探索文本与分子 2D/3D 信息的统一建模，推动三维结构显式纳入语言—结构联合表征，体现多模态科研基础模型从“对齐可用”向“结构与语义统一建模”演进<sup>39</sup>。

#### （四）科研智能体

**科研智能体**指能够在一定程度上自主执行科研活动的 AI 代理系统，其融合语言理解、规划决策、工具使用和环境交互等 AI 能力，旨在模拟或辅助研究人员的研究过程。科研智能体通常集成多个模型组件和工具模块，具备记忆、推理和行动能力，可围绕给定研究目标

e, 2023, 624(7990): 80-85.

<sup>38</sup> Liu Z, Li S, Luo Y, et al. Molca: Molecular graph-language modeling with cross-modal projector and uni-modal adapter[J]. arXiv preprint arXiv:2310.12798, 2023.

<sup>39</sup> Tang X, Tran A, Tan J, et al. MolLM: a unified language model for integrating biomedical text with 2D and 3D molecular representations[J]. Bioinformatics, 2024, 40(Supplement\_1): i357-i368.

生成一系列自主决策。根据自主程度的差异，科研智能体可视为“AI 科研助理”到“AI 科学家”的连续体：处于低级形态时，在人类指导下辅助完成部分任务；处于高级形态时，能在极少人类干预下自主循环执行完整的科研实验流程。科研智能体既可以是单一代理，也可以是多个专长各异的代理构成的协作体。

构建科研智能体需要在科研大模型的基础上融合多方面的关键技术。一是自动推理与工具使用，通过显式推理过程引导与工具调用机制，使大模型能够形成可检查的推理步骤，并使用接口调用文献搜索、计算程序等执行操作。二是自动实验设计与执行，在自动化实验室提供开放接口的前提下，科研智能体可将规划好的实验步骤转化为实验指令并下达，并结合实验数据反馈调整后续计划。三是多模态交互，科研智能体需要处理文本、图像、谱图、分子结构等模态数据，以全面感知实验环境并呈现科研成果。四是多智能体协同，通过引入多个不同专长的科研智能体，可以分工协作解决跨学科的复杂课题，该技术也能提升系统的模块化水平和可扩展性。

近年来，科研智能体领域涌现出诸多探索性成果，既包括面向通用科研流程的综合性平台，也涵盖聚焦特定科研领域的专用智能体。2024 年，Sakana AI 研究团队提出了全自动开放式科研框架 AI Scientist，该系统实现了科研循环的全流程自动化：从提出假设、检索文献、编写代码、运行实验，到结果分析、撰写论文，以及自动进行同行评议反馈，这一过程可以开放循环反复迭代<sup>40</sup>。在随后的 AI Scientist v2 版

---

<sup>40</sup> Lu C, Lu C, Lange R T, et al. The ai scientist: Towards fully automated open-ended scientific discovery[J]. arXiv preprint arXiv:2408.06292, 2024.

本中，其通过智能体树搜索摆脱了对人工代码模板的依赖，显著提升了系统在不同科研领域的通用性。并首次实现全流程自动生成的论文在国际知名学术会议专题研讨会投稿中通过同行评审（评审结果达到录用阈值），但该稿件按实验设计在评审完成后撤回，未进入最终正式录用与发表流程<sup>41</sup>。2025 年，谷歌开发的 AI Co-Scientist 系统定位为“虚拟科研合作者”，可协助生成假设、综述文献和设计实验方案，在生物医药研究的试点中已取得初步成效<sup>42</sup>。在化学材料领域，面向化学合成和材料设计的 ChemCrow、PolySea 等成果相继涌现，ChemCrow 能自主规划并执行有机合成路线、材料性质计算等复杂化学任务<sup>43</sup>；PolySea 则能根据目标性能要求生成全新的聚合物结构，并验证其可行性<sup>44</sup>。在社会学领域，研究者利用多智能体建模探索人类行为和社会运行的规律，如清华大学提出具有人类特征、由 LLM 赋能的智能体 EconAgent，用于宏观经济模拟<sup>45</sup>；该团队还推出大规模社会仿真系统 AgentSociety 1.0，其集成 1 万个智能体和 500 万次互动行为，用于研究舆论极化、谣言扩散与公共政策干预等议题<sup>46</sup>。

<sup>41</sup> Yamada Y, Lange R T, Lu C, et al. The ai scientist-v2: Workshop-level automated scientific discovery via agentic tree search[J]. arXiv preprint arXiv:2504.08066, 2025.

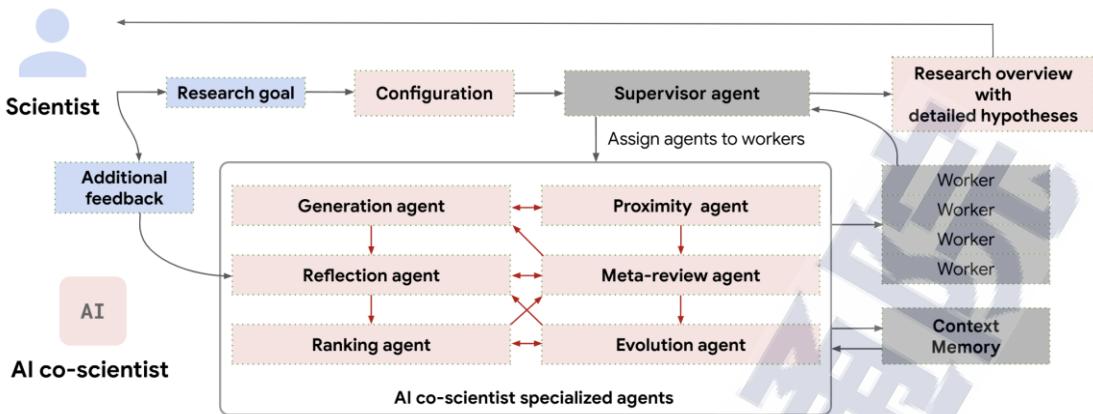
<sup>42</sup> Gottweis J, Weng W H, Daryin A, et al. Towards an AI co-scientist[J]. arXiv preprint arXiv:2502.18864, 2025.

<sup>43</sup> M. Bran A, Cox S, Schilter O, et al. Augmenting large language models with chemistry tools[J]. Nature Machine Intelligence, 2024, 6(5): 525-535.

<sup>44</sup> Qiu H, Zhao J, Jing E, et al. Introducing PolySea: An LLM-Based Polymer Smart Evolution Agent[J]. 2025.

<sup>45</sup> Li N, Gao C, Li M, et al. Econagent: large language model-empowered agents for simulating macroeconomic activities[C]//Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers). 2024: 15523-15536.

<sup>46</sup> Piao J, Yan Y, Zhang J, et al. Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society[J]. arXiv preprint arXiv:2502.08691, 2025.



来源：Towards an AI co-scientist

图 4 谷歌 AI Co-Scientist 多智能体架构设计示意图

## （五）自动化实验室

**自动化实验室**是指高度依赖机器人、自动控制和信息化系统来执行科研实验的实验环境，又称“自驱动实验室”、“智能实验室”。其侧重物理实验执行层面的自动化和智能化，样品准备、仪器操作、数据采集、初步分析等实验过程均由机器自动完成，并与上层 AI 决策系统无缝对接，可最大程度减少人工干预。自动化实验室可以显著提高实验的效率、一致性和重现性，为实现科研智能体的物理执行奠定基础。

构建自动化实验室依赖多学科技术的融合创新，包括自动化硬件系统、实验调度管理系统及数据管理平台等。自动化硬件系统主要包括固定式机械臂、移动机器人及各类具有自动进样、自动测量功能的实验装置，可最大限度赋予实验室执行复杂操作的能力，确保实验的高精度、高通量和可重复性。实验调度管理系统包括实验流程描述语言、任务规划算法和调度控制软件等，能将科研任务转化为具体实验步骤，并在时间和资源维度对多实验并行过程进行优化调度。数据管

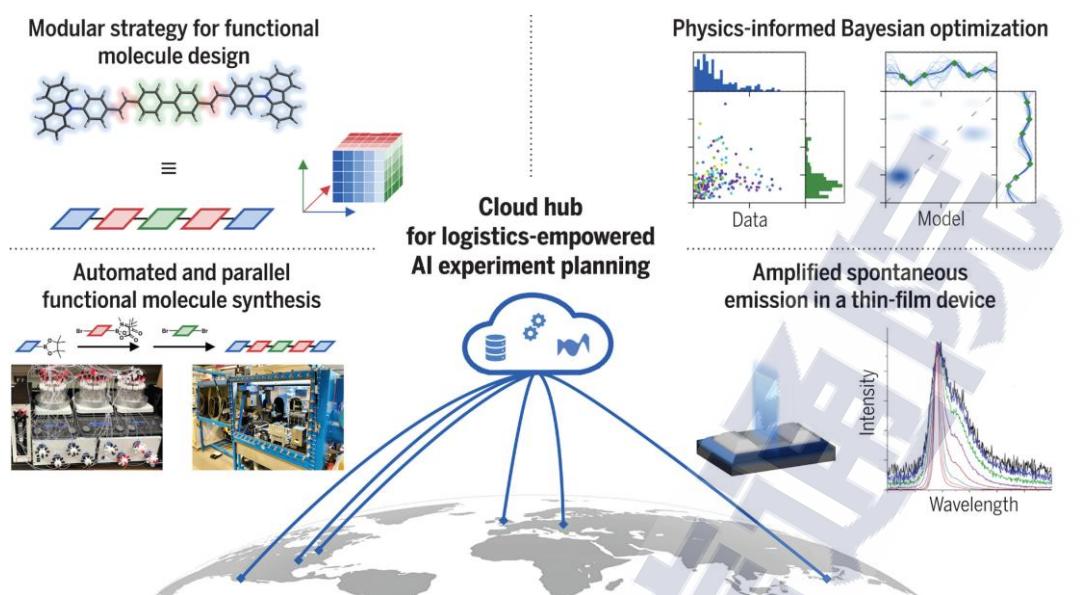
理平台需支持标准化的数据格式和数据库、自动记录和归档功能、开放共享与远程协同机制，确保实验数据可被及时分析和长期保存。

2020 年前后，全球涌现出一批标志性自动化实验室系统，并实现了从单个实验室到云端协同的跃迁。2020 年，英国利物浦大学研制出移动机器人化学家，其基于移动台和机械臂可以自主开展化学实验<sup>47</sup>。2021 年，中国科学技术大学研制出人工智能化学家，系统整合“化学大脑”（机器学习+贝叶斯优化+化学文献）、移动机器人和化学工作站，可实现从文献阅读、实验设计到合成测试的全流程自主执行<sup>48</sup>。2023 年，美国伯克利实验室的 A-Lab 平台构建了全自动化的材料合成与测试闭环系统，进一步整合主动学习、机器学习、文献知识、计算模拟与多机器人粉体合成线，验证“计算—实验”双向闭环范式。2024 年，加拿大多伦多大学联合全球 5 个实验室，将自动化实验室从单点扩展为“跨时区、跨设施、云端协同”的协作网络，实现分布式异步的材料设计—制备—测试—分析循环，并验证其有效性，为远程实验与实验资源云化奠定了基础<sup>49</sup>。

<sup>47</sup> Burger B, Maffettone P M, Gusev V V, et al. A mobile robotic chemist[J]. Nature, 2020, 583(7815): 237-241.

<sup>48</sup> Zhu Q, Zhang F, Huang Y, et al. An all-round AI-Chemist with a scientific mind[J]. National Science Review, 2022, 9(10): nwac190.

<sup>49</sup> Strieth-Kalthoff F, Hao H, Rathore V, et al. Delocalized, asynchronous, closed-loop discovery of organic laser emitters[J]. Science, 2024, 384(6697): eadk9227.



来源：Delocalized, asynchronous, closed-loop discovery of organic laser emitters

图 5 利用分布式实验室发现新材料的流程示意图

## 五、科研智能典型应用

科研智能正在引发一场深刻的科研范式变革。本章旨在系统地梳理近年来科研智能在“基础科学”和“产业研发”关键领域中的应用现状及成效。

### （一）驱动基础科学突破

在基础科学领域，AI 正帮助科学家应对三大经典挑战：高维“组合爆炸”的搜索空间、复杂系统的非线性模拟，以及海量实验数据的模式挖掘。AI 不仅在加速计算，更在提供一种全新的、数据驱动的科学“直觉”，帮助形成新的假说。本报告参考联合国教科文组织对基础科学的定义选取有代表性的学科进行分析<sup>50</sup>。

#### 1. 生命科学

生命科学正从“数据获取驱动”转向“数据理解与知识生成驱动”，

<sup>50</sup> UNESCO. Basic sciences [EB/OL]. [2026-01-11]. <https://www.unesco.org/en/basic-sciences>.

但这一转变被三类内生瓶颈牵制，三者构成“数据→机制→验证”的因果链，并形成反馈闭环。**一是数据理解的复杂性瓶颈。**单细胞与空间多组学将研究推向高维、多模态、强异质数据形态；空间组学甚至可在单次实验中产生 TB 级数据，显著抬升处理与分析门槛<sup>51</sup>。在跨实验室、跨队列整合时，批次效应虽可缓解但难以彻底消除，且在元信息不充分或混杂因素存在时可能引入偏差甚至伪相关。因此，数据越多并不自动转化为更可靠的知识，反而更容易让结论停留在相关性层面，难以稳定沉淀为可解释、可迁移的推断。**二是系统机制的黑箱挑战。**很多时候难点不在于找不到关联，而在于解释不清机制。以人类遗传学为例，全基因组关联研究（GWAS）已发现大量与疾病和性状相关的变异位点，但把这些统计关联变成可检验、可迁移的机制解释仍然困难。一个重要原因是：许多相关变异位于不编码蛋白质的非编码区，而这些位点在细胞类型分辨率上的功能注释仍不足。同时，从基因组结构看，蛋白编码区约占 2%，其余约 98% 为非编码区，包含大量调控信息和疾病相关变异，但“非编码信息如何被读取并最终影响表型”的机制链条仍难系统解释。**三是实验验证的“效率鸿沟”。**多组学与 AI 让候选机制、靶点和分子可以批量生成，但实验验证受限于周期长、流程不统一以及数据/元数据不完备，往往出现“候选很多、证据不足、难以收敛”的情况。与此同时，可重复性压力突出：PLOS Biology 期刊的国际调查显示，72% 的生物医学研究者认为领域存在可重复性危机；其中“发表压力”被认为是最主要诱因之一，62%

<sup>51</sup> Bressan D, Battistoni G, Hannon G J. The dawn of spatial omics[J]. Science, 2023, 381(6657): eabq4964.

的受访者认为其“总是或经常”导致不可重复<sup>52</sup>。验证跟不上会反过来减慢高置信知识与机制证据的增长，使数据整合更难校准、机制解释更难固化，从而让三类瓶颈相互强化。

近年来，AI 赋能生命科学的进展可概括为两类更具代表性的方向。在方向一中，以谷歌 DeepMind 为代表，沿“结构与互作理解→变异数量评估→调控预测”的链条，逐步推进对生命机制的系统解析：结构与互作提供分子层面的三维载体与作用界面，变异数量将基因变异映射到分子功能风险并支持优先级排序，调控预测则进一步面向非编码区，把调控变化与功能组学表征连接起来，补齐非编码变异解释的关键环节。结构与互作层面，AlphaFold 3（2024 年）将建模对象从单一蛋白拓展到多类分子复合体，可预测包含蛋白、核酸、小分子、离子等在内的复合体联合三维结构，为分子互作研究提供更通用的结构化支撑。与此同时，结构数据资源的“平台化沉淀”与模型能力迭代相互促进。以 AlphaFold 蛋白结构数据库（AFDB）为例，该库由欧洲分子生物学实验室—欧洲生物信息研究所（EMBL-EBI）与谷歌 DeepMind 联合建设，面向科研人员开放提供高精度的蛋白结构预测数据，目前数据库条目规模已达 2 亿+，并被集成到主流数据库、可视化平台与分析流程中<sup>53</sup>；公开信息显示，AFDB 已被全球 300 万+研究人员使用，覆盖 190+个国家/地区，体现其持续嵌入生命科学研究的日常工作流<sup>54</sup>。变异数量层面，AlphaMissense（2023 年）对所有可

<sup>52</sup> Cobey K D, Ebrahimzadeh S, Page M J, et al. Biomedical researchers' perspectives on the reproducibility of research[J]. PLoS biology, 2024, 22(11): e3002870.

<sup>53</sup> Google DeepMind, European Bioinformatics Institute. AlphaFold Protein Structure Database [EB/OL]. [2026-01-03]. <https://alphafold.ebi.ac.uk/>.

<sup>54</sup> Google DeepMind. AlphaFold: Five years of impact[EB/OL]. (2025-11-25)[2026-01-03]. <https://deepmin>

能的约 7100 万个错义变异给出“可能致病/可能良性”的效应倾向预测，可用于遗传变异解释与优先级排序<sup>55</sup>。调控层面，AlphaGenome（2025 年）进一步面向非编码调控区，以最长约 1 Mb 的 DNA 序列为输入，在单碱基分辨率上预测数千种分子属性/功能组学轨道，用于评估非编码变异的潜在分子影响<sup>56</sup>。

在方向二中，以华盛顿大学戴维·贝克（David Baker）教授团队为代表，推动生成式模型从预测走向设计，提升生物分子设计与验证的效率。贝克团队研发的 RFdiffusion（2023 年）提出了基于扩散模型的从头蛋白设计通用框架，并通过对大量设计产物的实验表征与结构验证，证明该路线具备可复用的工程化基础。在扩散模型路线之外，蛋白语言模型也开始展示“生成—合成—功能验证”的能力，ESM3（2025 年）在提示驱动下生成荧光蛋白序列，研究团队合成分后获得明亮荧光蛋白，且该蛋白与已知荧光蛋白的序列一致性仅约 58%，论文估算其相当于“模拟了约 5 亿年的进化距离”，为“生成式模型可探索远离天然序列空间并产生可功能验证分子”提供了代表性证据。在此基础上，贝克团队进一步将生成能力拓展到更具挑战性的目标场景：2025 年的研究展示了生成式 AI 可面向固有无序蛋白/无序区域设计高精度结合蛋白，为处理“难以用稳定结构描述”的靶点提供了新的设计路径<sup>57</sup>。进一步地，在抗体设计场景中，该团队将模型的定向优

[d.google/blog/alphafold-five-years-of-impact/](https://d.google/blog/alphafold-five-years-of-impact/)

<sup>55</sup> Cheng J, Novati G, Pan J, et al. Accurate proteome-wide missense variant effect prediction with Alpha Missense[J]. Science, 2023, 381(6664): eadg7492.

<sup>56</sup> Avsec Ž, Latysheva N, Cheng J, et al. AlphaGenome: advancing regulatory variant effect prediction with a unified DNA sequence model[J]. bioRxiv, 2025: 2025.06. 25.661532.

<sup>57</sup> Liu C, Wu K, Choi H, et al. Diffusing protein binders to intrinsically disordered proteins[J]. Nature, 2025, 644(8077): 809-817.

化与酵母展示筛选结合，实现按指定表位生成抗体，并通过结构数据验证其 6 个互补决定区（CDR）环的构象设计达到原子级精度，体现了“生成—筛选—结构验证”的闭环方法在复杂生物分子设计中的可行性与可验证性<sup>58</sup>。同时，产业界也在推进“按需设计结合体”的工程化落地。以 DeepMind 的 AlphaProteo（2024 年）为例，该工具面向蛋白质结合体设计，旨在针对给定靶点蛋白生成新的高亲和力结合蛋白。公开结果显示，其在多靶点测试中报告更高实验成功率，并给出相对既有方法 3–300 倍的结合亲和力提升，为“生成式模型可用于获得高亲和结合分子”的工程潜力提供了更具量化特征的证据支撑<sup>59</sup>。

## 2. 化学

化学作为一门“创造新物质”的科学，其研发范式长期受到三大内生挑战的牵制。**一是搜索的组合爆炸。**化学空间（理论上可能的分子、反应等的总和）与反应空间规模及其庞大。例如，在材料发现中，潜在可组合的材料候选空间被估算可超过  $10^{60}$  种化合物量级，这使得依靠经验与低通量试错很难实现系统覆盖与高效探索。**二是合成的路径鸿沟。**即使在理论上给出目标分子结构，“如何可行、经济且可靠地把它做出来”仍是核心瓶颈。逆向合成需要在巨大分支的可能性图中搜索路径，更关键的是候选路线必须满足起始物料可得性与成本等供应链约束，而这些信息在规划阶段往往难以精确量化；同时还必须满足环境-健康-安全（EHS）与过程安全等约束，否则容易出现“纸

<sup>58</sup> Bennett N R, Watson J L, Ragotte R J, et al. Atomically accurate de novo design of antibodies with RFdiffusion[J]. Nature, 2025: 1-11.

<sup>59</sup> Zambaldi V, La D, Chu A E, et al. De novo design of high-affinity protein binders with AlphaProteo [J]. arXiv preprint arXiv:2409.08022, 2024.

面可行但难以落地实施”的鸿沟。**三是机理的认知黑箱。**在不少（尤其涉及催化循环与多步耦合的）反应体系中，机理解析成本高且证据链不完备，而机理理解又对新催化剂与新反应性设计至关重要，导致优化与创新仍大量依赖经验迭代与试错。

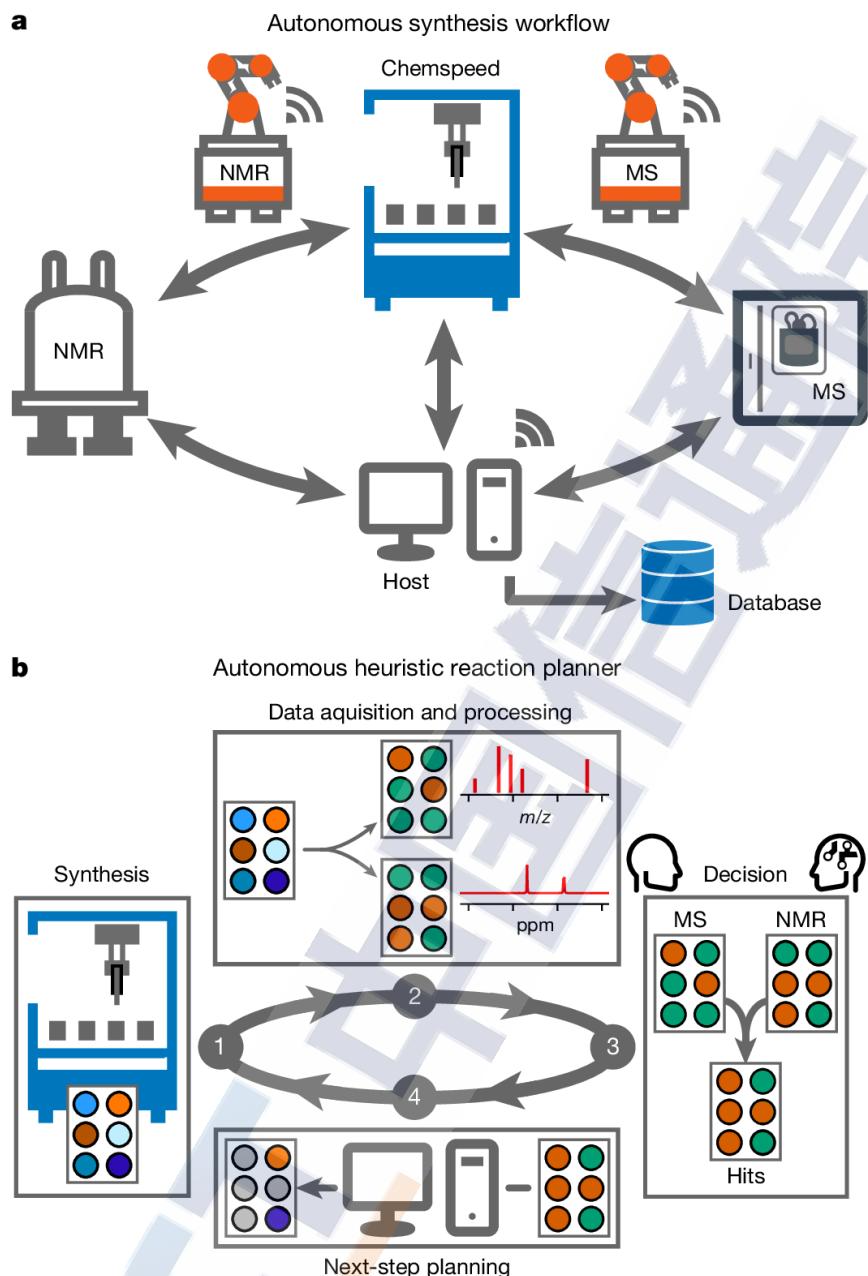
人工智能与自动化、机器人平台加速融合，推动化学研究从“可计算”走向“可执行、可验证、可复现”的闭环范式。相关进展可概括为两条主线：其一，依托自动化/机器人打通“设计—执行—分析—学习”的物理闭环，提升实验流程的并行化与执行能力，解决“做得到”；其二，通过指标体系与复现导向的工程化抽象，并引入守恒/可行性约束，增强闭环的可度量、可审计与可迁移性，提升“算得准、判得准、做得成”，抑制误差传播。一方面，自动化实验室正在打通“从设计到执行”的全流程闭环，使 AI 从虚拟推理进一步走向实验调度与执行。典型代表之一是 ChemCrow（2024 年），其将 GPT-4 与 18 个化学专用工具集成，使智能体能够在工具链支持下进行多步推理与操作编排，并在实验任务中实现自主规划并执行合成，体现了“工具使用+多步推理”对化学任务流自动化与流程化管理的价值。同期利物浦大学（2024 年）展示了基于移动机器人的模块化自主平台，移动机器人能够将合成环节与多种分析模块连接，并能在与人类研究人员共享设备的条件下运行；更重要的是，该工作明确强调“正交表征（多手段交叉验证）对于降低误判与不确定性至关重要”，并展示了在不改变既有仪器的情况下接入台式核磁共振与超高效液相色谱-质

谱联用等设备，从而降低自治能力嵌入常规化学实验室的门槛<sup>60</sup>。在材料合成与表征闭环方向，系统瓶颈往往不在“能否自动做”，而在“表征—判读—标准化”的可靠性：A-Lab（2023 年）虽展示了高吞吐闭环能力，但后续外部分析对其“新材料/成功判定标准”提出质疑，并指出仅凭拟合优度等单一统计指标可能导致误判，提示判读可靠性将直接决定闭环上限<sup>61</sup>。针对上述问题，近两年的一些工作开始在系统层面强化在线监测、数据分析与人机协同决策，使闭环不仅追求更快迭代，也更强调过程可审计与结果可验证。以 2025 年的一项材料体系研究为例，研究者引入 AI 顾问进行实时进度监测与数据分析，并在 64 次实验迭代内实现性能快速优化，最终揭示了此前未报道的聚合物多晶型，体现了闭环在增强判读链路后可提升发现的可检验性<sup>62</sup>。

<sup>60</sup> Dai T, Vijayakrishnan S, Szczypinski F T, et al. Autonomous mobile robots for exploratory synthetic chemistry[J]. Nature, 2024, 635(8040): 890-897.

<sup>61</sup> Peplow M. Robot chemist sparks row with claim it created new materials[J]. Nature, 2023.

<sup>62</sup> Dai Y, Chan H, Vriza A, et al. Adaptive AI decision interface for autonomous electronic material discovery[J]. Nature Chemical Engineering, 2025: 1-11.



来源：Autonomous mobile robots for exploratory synthetic chemistry

图 6 模块化机器人工作流程和启发式反应规划器示意图

另一方面，为补齐闭环中的可信设计与可信判读，模型与规划方法正系统性引入化学基本约束与面向落地的可行性约束，推动闭环从“跑通流程”走向“结果可信”。在反应建模层面，研究正在从“数据驱动的经验拟合”转向“守恒与机理一致性约束”的可控学习范式，即在反应表示与生成过程中显式引入化学基本约束，以降低不符合化

学规律的输出风险并提升可检验性。典型工作如 2025 年的 FlowER，通过在反应表示与生成中显式施加质量守恒，并在表示层面同时约束质量与电子守恒，从源头缓解反应预测中的幻觉式错误模式，增强结果的化学一致性与可解释性<sup>63</sup>。在逆合成规划方向，研究重点也由“仅追求搜索效率与路径长度”逐步转向“面向真实合成可行性的约束建模与评价体系完善”。一方面，算法开始将“指定起始物料”等现实约束纳入规划问题设定，并通过搜索策略确保约束可满足（如 DESP<sup>64</sup>），使逆合成从“理论可达”进一步迈向“路径可执行”。另一方面，路线级评估指标从算法友好的内部一致性度量，延伸到对化学合理性/可行性的刻画，例如 Retro-BLEU（2024 年）用于评估合成路线的“可行性/可信度”，并能够区分更可信的可行路线与不合理路线，为闭环系统在多候选路线中筛选“更可信、更可做”的方案提供依据<sup>65</sup>。

### 3. 地球与空间科学

地球与空间科学的核心挑战可以概括为相互耦合的“三个瓶颈”：系统本体的复杂性、观测体系的约束、以及模型与数据的工程化与可信性。首先，研究对象普遍呈现强非线性、跨尺度耦合与显著的可预报性边界。以大气（天气）系统为代表，其混沌特征叠加多尺度强耦合过程，使得数值预报与地球系统模拟必须在更高分辨率、更丰富物

<sup>63</sup> Joung J F, Fong M H, Caselli N, et al. Electron flow matching for generative reaction mechanism prediction[J]. Nature, 2025, 645(8079): 115-123.

<sup>64</sup> Yu K, Roh J, Li Z, et al. Double-ended synthesis planning with goal-constrained bidirectional search [J]. Advances in Neural Information Processing Systems, 2024, 37: 112919-112949.

<sup>65</sup> Li J, Fang L, Lou J G. Retro-BLEU: quantifying chemical plausibility of retrosynthesis routes through reaction template sequence analysis[J]. Digital Discovery, 2024, 3(3): 482-490.

理过程表征以及集合化不确定性量化等方面持续加码，从而对计算资源、时效与稳定性提出高强度要求；同时，临界阈值、关键反馈及其潜在级联效应的机理与约束仍存在显著不确定性，导致在决策所需的时空尺度上难以给出高置信度的风险刻画。**其次，观测体系本身构成预测与认知能力的上限约束：**观测网络在时空覆盖、关键变量可观测性、长期连续记录与跨代际任务接续方面仍存在短板，而“观测—数据同化—初值”链条的误差会快速放大并主导后续预报偏差；在空间天气等方向，高影响事件样本相对稀缺且关键观测更依赖连续运行，使这一约束更为突出。**再次，领域在“数据规模—数据可用性—模型可信性”之间面临结构性张力：**新一代观测与模拟持续产出海量、质量参差且跨源异构的数据（噪声、缺测、系统性偏差与分辨率不匹配并存），将其沉淀为可复用、可追溯、可直接支撑分析与训练的数据资产，往往需要质量控制、预处理、配准与融合、元数据规范化以及样本构建（含弱标注/自动标注）等复杂流程，工程成本高、周期长；与此同时，端到端数据驱动模型即便在部分指标上取得突破，也可能在复杂耦合物理规律面前出现物理一致性不足、过程不可解释、概率校准不稳定等“物理失真”问题，尤其在极端事件与分布漂移情景下会放大为业务与决策风险。

接下来，分别从地球科学和空间科学两个维度介绍人工智能的赋能进展。**地球科学领域，AI 赋能主要体现在预报模拟与地球观测两条主线。气象预报模拟正在从确定性走向概率集合，并进入业务/准业务运行体系。**以谷歌 GenCast（2025 年）为例，其公开说明可生成

50 个或更多的集合预报，并给出单个集合成员 15 天预报约 8 分钟且可并行生成的效率口径，显著降低集合预报的计算门槛。另一条技术路线突破是混合地球系统建模，NeuralGCM（2024 年）将可微动力学求解器与机器学习组件结合，试图同时兼顾预报技巧、集合预报与长期稳定性，凸显物理一致性与稳定性仍是关键攻关方向。同期，国际顶级业务预报中心已将 AI 能力纳入业务运行体系。欧洲中期天气预报中心（ECMWF）于 2025 年上线其 AI 预报系统 AIFS<sup>66</sup>，并推出集合预报版本（AIFS ENS）<sup>67</sup>。ECMWF 同时明确：集合预报的初始条件仍主要依赖物理同化系统生成。这表明其工程化路径并非以 AI 完全替代既有链路，而是采取以 AI 增强预报能力、并与既有业务流程耦合迭代的渐进式演进路线。我国方面，公开信息显示，中国气象部门已对国产 AI 预报模型开展准业务评估与试运行，并通过示范计划等机制形成阶段性评估结论、探索业务准入与本地化应用路径<sup>68</sup>。但独立评估也表明，在破纪录极端事件刻画上，传统数值预报总体更稳健，AI 模型存在系统性低估极端事件频次与强度的倾向，极端尾部风险与可信表达仍是关键短板<sup>69</sup>。在此基础上，一个更具长期意义的共同趋势正在显现，预报与观测都在走向基础模型化。在预报侧，微软 Aurora（2025 年）通过大规模预训练与高效微调，将天气、空气质

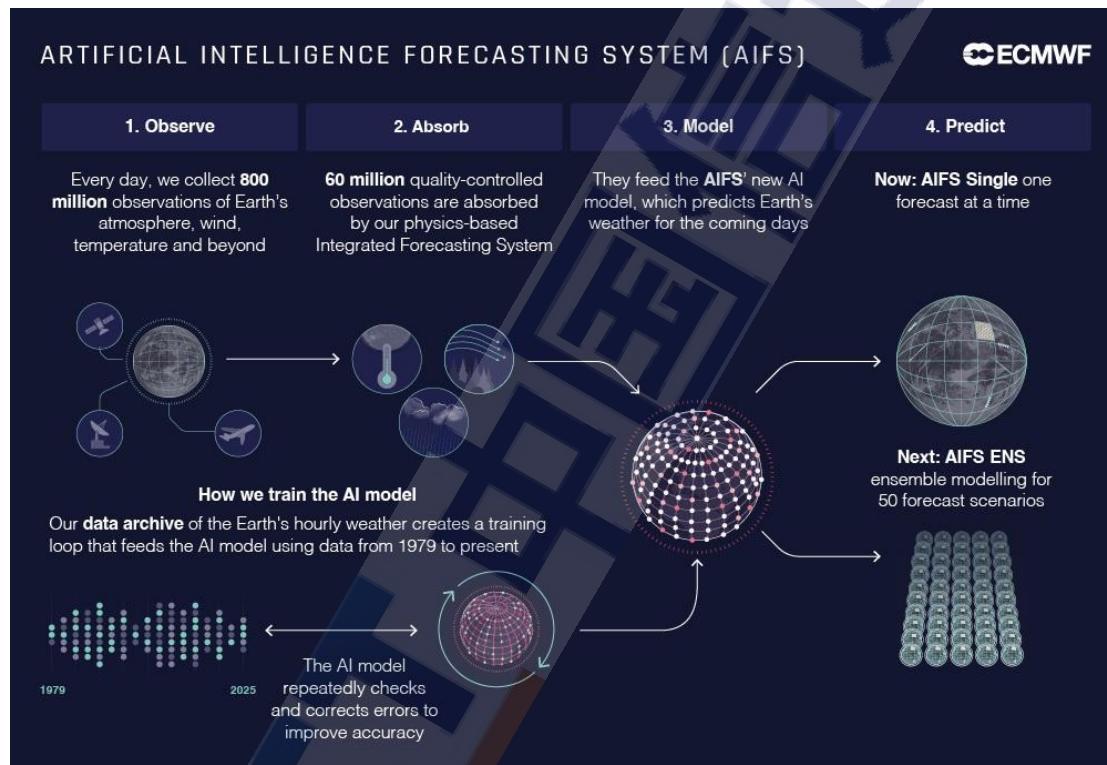
<sup>66</sup> European Centre for Medium-Range Weather Forecasts (ECMWF). ECMWF's AI forecasts become operational[EB/OL]. (2025-02-25)[2025-12-30]. <https://www.ecmwf.int/en/about/media-centre/news/2025/ecmwfs-ai-forecasts-become-operational>

<sup>67</sup> European Centre for Medium-Range Weather Forecasts (ECMWF). ECMWF's ensemble AI forecasts become operational[EB/OL]. (2025-07-01)[2025-12-30]. <https://www.ecmwf.int/en/about/media-centre/news/2025/ecmwfs-ensemble-ai-forecasts-become-operational>

<sup>68</sup> 中国气象局. 中国气象局推动人工智能天气预报模型创新融合应用[EB/OL]. (2025-09-10)[2025-12-30]. [https://www.cma.gov.cn/2011xwzx/2011xmtjj/202509/t20250910\\_7324169.html](https://www.cma.gov.cn/2011xwzx/2011xmtjj/202509/t20250910_7324169.html)

<sup>69</sup> Zhang Z, Fischer E, Zscheischler J, et al. Numerical models outperform AI weather forecasts of record-breaking extremes[J]. arXiv preprint arXiv:2508.15724, 2025.

量、海浪等多类地球系统预测任务纳入统一框架，体现出从单点模型能力走向可复用预测底座的趋势<sup>70</sup>。在感知侧，地球观测基础模型（如 Prithvi-EO-2.0）通过大规模预训练形成可迁移表征，并以轻量适配方式支撑洪水、生态、农业、城市等任务，推动遥感能力通用化，降低下游任务从零训练的成本，体现出“可复用底座+统一评测”的发展方向<sup>71</sup>。



来源：欧洲中期天气预报中心（ECMWF）网站

图 7 ECMWF AI 预报系统 AIFS 工作流程示意图

空间科学领域，AI 的赋能主要体现在两类高价值链路，一是天文巡天中的瞬变事件发现与告警处置，二是日球物理（空间天气）中的提前预警与形态预测。在天文巡天场景中，AI 正由专用分类器升

<sup>70</sup> Bodnar C, Bruinsma W P, Lucic A, et al. A foundation model for the Earth system[J]. Nature, 2025: 1-8.

<sup>71</sup> Szwarcman D, Roy S, Fraccaro P, et al. Prithvi-eo-2.0: A versatile multi-temporal foundation model for earth observation applications[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025.

级为更通用的发现助手，其核心意义在于显著降低“海量告警→少量真信号”的筛选门槛：牛津大学的研究表明（2025 年），仅用“15 个示例图像+简要指令”，通用大模型即可在“真实宇宙事件与成像伪影”分类上达到约 93% 准确率，并能对每次分类给出通俗解释，从而提升透明度与可用性<sup>72</sup>。同期，牛津大学针对超新星告警筛选的成果进一步给出可量化的生产力证据，称新工具可将天文学家相关工作量降低约 85%，凸显其在高通量数据流中的降本增效潜力<sup>73</sup>。在空间天气方向，基础模型思路开始向日球物理延伸：NASA 介绍的 Surya 模型（2025 年）提出可对太阳耀斑进行提前约两小时的可视化预测，并报告在既有基准上取得约 16% 的改进<sup>74</sup>；IBM 同日发布信息强调其可生成高分辨率图像，用于预测耀斑可能发生的位置与形态、提前量可达 2 小时<sup>75</sup>。

#### 4. 数学

数学作为一门古老而根基深厚的学科，其当前的发展面临着一系列深刻的挑战。**一是知识规模爆炸与学科高度分化**。根据美国数学学会（AMS）的最新统计，权威数据库 MathSciNet 收录的文献记录目前已超过 400 万条，仅 2024 年单年新增文献就达到近 13 万篇，且近年持续保持这一高位增长，这一增速远超任何单个研究者的全量追踪

<sup>72</sup> Stoppa F, Bulmus T, Bloemen S, et al. Textual interpretation of transient image classifications from large language models[J]. Nature Astronomy, 2025: 1-10.

<sup>73</sup> Stevance H F, Smith K W, Smartt S J, et al. The ATLAS Virtual Research Assistant[J]. arXiv preprint arXiv:2506.09778, 2025.

<sup>74</sup> NASA Science. NASA, IBM's 'Hot' New AI Model Unlocks Secrets of Sun[EB/OL]. (2025-08-20)[2025-12-30]. <https://science.nasa.gov/science-research/artificial-intelligence-model-heliophysics/>

<sup>75</sup> IBM. IBM and NASA Release Groundbreaking Open-Source AI Model on Hugging Face to Predict Solar Weather and Help Protect Critical Technology[EB/OL]. (2025-08-20)[2025-12-30]. <https://newsroom.ibm.com/2025-08-20-ibm-and-nasa-release-groundbreaking-open-source-ai-model-on-hugging-face-to-predict-solar-weather-and-help-protect-critical-technology>

能力<sup>76</sup>。这导致了知识碎片化与信息不对称的风险，研究者若难以把握全局，可能导致重复劳动。此外，学科分化使得跨领域深层联系的识别往往依赖少数具备跨域背景的研究者、综述机制与学术共同体网络，使得系统性的跨领域知识整合难度不断上升。**二是证明复杂性与可验证性压力。**许多前沿成果的证明往往跨越数篇论文、篇幅浩繁，对同行评议与学术共同体的“可验证性能力”提出挑战。以近年来几何朗兰兹相关研究为例，其核心工作分布在 5 篇论文中，总长超过 800 页，并建立在长期技术积累之上。针对这类超长证明，已有研究指出：尽管理想状态下可通过逐行审阅来核对正确性，但在现实学术实践中，完全逐行核查往往难以实现，从而引发关于“可验证性边界”与审稿机制的持续讨论<sup>77</sup>，这引发了界内对“可验证性”现实边界的讨论。为应对此瓶颈，部分研究者开始利用 Lean、Isabelle 等工具推动“形式化转向”，但目前正如相关领域专家所言，利用证明助手完全验证“全新且高难度工作”的案例在现阶段仍然较少，尚未成为普遍的工作流。**三是结构探索的“组合爆炸”与系统性覆盖不足。**在数论、组合等领域，数学家面对的是呈现指数级增长的搜索空间。以 2024 年的 FunSearch 工作为例，其在解决帽子集（Cap Set）问题时，作者明确指出搜索空间会迅速膨胀至极其巨大的量级（例如  $n=8$  时可达约  $3^{1600}$  的规模）<sup>78</sup>，使得仅依赖人类直觉的探索难以实现系统性覆盖。

<sup>76</sup> AMERICAN MATHEMATICAL SOCIETY. Mathematical Reviews 2024 Annual Report[R/OL]. (2025-01)[2025-12-25]. <https://www.ams.org/publications/math-reviews/MR-Annual-Report-2024.pdf>.

<sup>77</sup> Greiffenhagen C. Checking correctness in mathematical peer review[J]. Social studies of science, 2024, 54(2): 184-209.

<sup>78</sup> Romera-Paredes B, Barekatain M, Novikov A, et al. Mathematical discoveries from program search with large language models[J]. Nature, 2024, 625(7995): 468-475.

即便是旨在自动化发现公式的拉马努金机器 (Ramanujan Machine) 项目，其团队在 2023 年的后续研究也指出传统策略仍可能依赖穷举式搜索，从而在覆盖广阔候选空间时存在天然局限<sup>79</sup>。因此，在缺乏更强的自动搜索与验证框架支持时，潜在的新结构与新猜想可能因启发式偏置与覆盖不足而被延后发现甚至长期遗漏。

近年来，AI 对数学研究的推动可概括为三个方向：形式化与验证降本增效、高难度推理与求解能力上限提升、以及结构与猜想的自动生成与探索。第一，在自动定理证明与形式化验证方面，研究重心正从概念验证转向可复现的开源基座与端到端系统。2023 年的 LeanDojo 将面向 Lean 的工具链、数据与基准以开源方式系统化，并将“前提选择”明确为定理证明的关键瓶颈之一；其提出的检索增强证明器 ReProver 通过从大型数学库检索可用前提，并结合大语言模型进行证明搜索，降低了实验复现门槛并提升了基线效果<sup>80</sup>。随后，形式化证明进一步呈现“合成冷启动数据+强化学习/搜索策略”的训练范式，例如 DeepSeek-Prover V2 (2025 年) 在 MiniF2F 测试集上报告 88.9% 的通过率，并披露了合成冷启动数据构建与后续强化学习阶段的训练流程<sup>81</sup>。同期也出现了更高通过率的端到端系统（如 Delta Prover 在 MiniF2F 测试集上报告 95.9% 通过率），显示“检索/分解/反思/搜索”与大模型训练的组合正在逼近可用性门槛<sup>82</sup>。需要强调的

<sup>79</sup> Razon O, Harris Y, Gottlieb S, et al. Automated search for conjectures on mathematical constants using analysis of integer sequences[C]//International Conference on Machine Learning. PMLR, 2023: 28809-28842.

<sup>80</sup> Yang K, Swope A, Gu A, et al. Leandojo: Theorem proving with retrieval-augmented language models [J]. Advances in Neural Information Processing Systems, 2023, 36: 21573-21612.

<sup>81</sup> Ren Z Z, Shao Z, Song J, et al. Deepseek-prover-v2: Advancing formal mathematical reasoning via reinforcement learning for subgoal decomposition[J]. arXiv preprint arXiv:2504.21801, 2025.

<sup>82</sup> Zhou Y, Zhao J, Zhang Y, et al. Solving formal math problems by decomposition and iterative reflecti

是，这一方向的进步不仅体现为分数提升，也依赖评测口径的可比性：例如应明确题集版本与划分、统计指标（通过率或 Pass@k）、采样预算与搜索上限、检索库版本、以及是否端到端覆盖完整证明链路，避免不同设置下的分数被直接横向比较。

**第二，在竞赛级难题求解方面**，在明确评测流程与系统设置的条件下，AI 在国际数学奥林匹克竞赛（IMO）中已达到可量化的高水平。2024 年，DeepMind 报告 AlphaProof 与 AlphaGeometry 2 在 IMO 2024 中解出 4/6 题、得 28/42 分，达到银牌区间，同时指出该流程仍依赖专家将题目手工翻译为形式化语言且计算耗时可达数天<sup>83</sup>。2025 年，DeepMind 报告 Gemini Deep Think 在自然语言端到端设置下解出 5/6 题、得 35/42 分，并称由 IMO 相关评审确认达到金牌分数线<sup>84</sup>。同年，OpenAI<sup>85</sup>和 DeepSeek<sup>86</sup>也分别宣布其大模型系统在公开披露的评测设置下（不使用工具或互联网、遵循竞赛时间及尝试次数）达到金牌分数线。

**第三，在结构发现与猜想生成方面**，AI 正从解题者向探索者与提出者扩展。一类路径是“形式化环境中的猜想生成管道”，例如 LeanConjecturer（2025 年）提出在 Lean 4 中自动生成大学层次猜想并

on[J]. arXiv preprint arXiv:2507.15225, 2025.

<sup>83</sup> AlphaProof and AlphaGeometry teams. AI achieves silver-medal standard solving International Mathematical Olympiad problems[EB/OL]. (2024-07-25)[2025-12-29]. <https://deepmind.google/blog/ai-solves-imo-problems-at-silver-medal-level/>

<sup>84</sup> LUONG T, LOCKHART E. Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the International Mathematical Olympiad[EB/OL]. (2025-07-21)[2025-12-29]. <https://deepmind.google/blog/advanced-version-of-gemini-with-deep-think-officially-achieves-gold-medal-standard-at-the-international-mathematical-olympiad/>

<sup>85</sup> aw31. OpenAI IMO 2025 Proofs[EB/OL]. (2025-07-18)[2025-12-29]. <https://github.com/aw31/openai-imo-2025-proofs/>

<sup>86</sup> Liu A, Mei A, Lin B, et al. Deepseek-v3. 2: Pushing the frontier of open large language models[J]. arXiv preprint arXiv:2512.02556, 2025.

进行筛选与迭代的流程，并报告了可量化的生成与过滤结果<sup>87</sup>。另一类路径是将大语言模型与可自动验证评估器结合，开展高通量探索，FunSearch（2023 年）在帽子集（cap set）等极值组合问题上，借助自动评估约束发现了超过既有最好结果的构造（如在  $n=8$  时找到规模为 512 的帽子集，优于此前公开的 496），表明“可自动验证+大规模搜索”的方式有望触及人类直觉不易覆盖的结构空间<sup>88</sup>。与此同时，面向数学常数的公式自动发现也在近三年取得明显进展：相关研究在继承早期“拉马努金机器”路线的基础上，提出了可用于统一与生成大规模公式族的新结构（如“保守矩阵场”），并在《美国国家科学院院刊》等渠道对其机理与适用范围进行了系统阐释<sup>89</sup>；此外，还出现了面向海量论文的工程化范式——通过“公式自动抽取—等价关系证明—知识统一”的流水线，从数十万篇 arXiv 论文中自动采集公式并证明它们的等价关系，从而提升跨论文知识关联发现的规模化能力<sup>90</sup>。

## （二）加速产业研发进程

如果说 AI 在基础科学中的应用是“发现不可能”，那么在产业研发中的应用就是“逆转不经济”。AI 正被用于解决产业界最头疼的“成本、周期、成功率”三大难题，重塑相关产业的竞争格局和价值链。本报告重点分析 AI 在医药、材料、半导体和先进制造等关键产

<sup>87</sup> Onda N, Kasaura K, Oriike Y, et al. LeanConjecturer: Automatic Generation of Mathematical Conjectures for Theorem Proving[J]. arXiv preprint arXiv:2506.22005, 2025.

<sup>88</sup> Romera-Paredes B, Barekatain M, Novikov A, et al. Mathematical discoveries from program search with large language models[J]. Nature, 2024, 625(7995): 468-475.

<sup>89</sup> Elimelech R, David O, De la Cruz Mengual C, et al. Algorithm-assisted discovery of an intrinsic order among mathematical constants[J]. Proceedings of the National Academy of Sciences, 2024, 121(25): e2321440121.

<sup>90</sup> Raz T, Shalyt M, Leibtag E, et al. From Euler to AI: Unifying Formulas for Mathematical Constants [J]. arXiv preprint arXiv:2502.17533, 2025.

业的研发赋能。

## 1. 医药研发

当前，医药研发产业面临的挑战可概括为研发效率持续下滑这一总体困境，以及导致这一困境的两类结构性原因。总体困境体现为反摩尔定律下的高成本与长周期。自 20 世纪 50 年代以来，医药行业面临严峻的投入产出比下降趋势。相关研究指出，在通胀调整口径下，每 10 亿美元研发投入所对应的获美国食品药品监督管理局（FDA）批准上市的新药数量，大约每 9 年减半，这一现象被称为“反摩尔定律”（Eroom's Law）<sup>91</sup>。从过程维度看，一款新药从概念发现到获批上市通常需要耗时 10–15 年。尽管不同机构的测算口径存在差异（如是否包含资本化成本、失败项目等），但行业主流引用的估算数据显示，单款新药的资本化研发成本多处于 20 多亿美元量级<sup>92</sup>。总体而言，投入持续攀升而产出效率下滑，构成了行业的基础性效率困境。关键原因之一是临床试验的转化困境。临床阶段往往是成本消耗最集中的环节之一。根据美国生物技术创新组织（BIO）对 2011–2020 年近万个药物研发项目的分析，从 I 期临床到最终获批的总体成功率仅约 7.9%，这意味着在该统计口径下，超过 90% 的项目无法成功上市。其中，II 期通常被认为是最关键、也最陡峭的瓶颈之一（在部分统计口径下通过率不足 30%）<sup>93</sup>。回顾性研究显示，临床试验失败的直接原因多为

<sup>91</sup> Scannell J W, Blanckley A, Boldon H, et al. Diagnosing the decline in pharmaceutical R&D efficiency [J]. Nature reviews Drug discovery, 2012, 11(3): 191-200.

<sup>92</sup> DiMasi J A, Grabowski H G, Hansen R W. Innovation in the pharmaceutical industry: new estimates of R&D costs[J]. Journal of health economics, 2016, 47: 20-33.

<sup>93</sup> Biotechnology Innovation Organization, Informa Pharma Intelligence, QLS Advisors. Clinical Development Success Rates and Contributing Factors 2011-2020 [R/OL]. (2021-02) [2025-12-25]. [https://go.bio.org/rs/490-EHZ-999/images/ClinicalDevelopmentSuccessRates2011\\_2020.pdf](https://go.bio.org/rs/490-EHZ-999/images/ClinicalDevelopmentSuccessRates2011_2020.pdf).

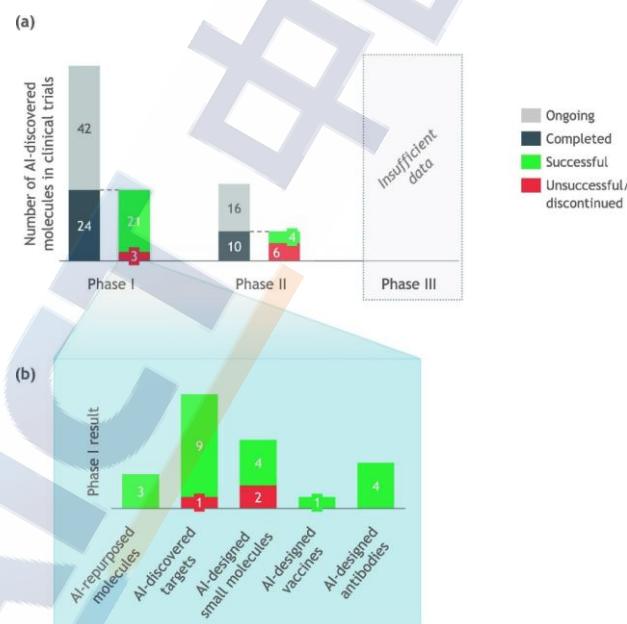
疗效不足或安全性问题，而在更深层面，临床前模型（细胞、动物模型等）对人体真实生物学反应的预测效度有限，被普遍视为造成转化落差的重要背景因素之一，从而形成难以跨越的“转化医学鸿沟”<sup>94</sup>。

**关键原因二是数据、科学与组织协同的范式困境。**一方面是数据与组织的割裂，研发、临床、基因组学等海量数据分散在不同组织和缺乏互操作性的遗留系统中，难以被有效整合利用；在医药企业内部，生物学家、化学家、临床医生与数据科学家之间长期存在的“知识与流程竖井”，进一步加剧了跨学科协同的难度。另一方面是科学复杂度的跃迁，全球药物管线研发重心正转移至肿瘤、免疫、神经科学及细胞与基因治疗等高复杂性领域，这些新兴方向往往面临疾病机制复杂、缺乏成熟模型及长期随访数据的挑战，使得传统研发范式越来越难以应对。

近年来，AI 正从分子发现、实验模式到临床开发三个关键维度，推动产业应对前述的效率与转化挑战。**一是前端提效：在早期发现与分子设计环节，AI 通过“数据驱动的靶点识别+生成式分子设计”实现提速与提质。**这也是目前公开可量化证据相对更充分、产业应用进展更快的方向之一。根据 2024 年针对“AI 原生”生物医药企业公开披露管线的统计研究，截至 2023 年底已有 75 个由 AI 设计的分子进入临床试验，其中 67 个仍处于在研状态，显示出近五年进入临床的数量增长较快。在已完成 I 期临床的 24 个药物样本中，有 21 个成功进入下一阶段，I 期过渡成功率达到 87.5%；作为对照，BIO 的行业

<sup>94</sup> Seyhan A A. Lost in translation: the valley of death across preclinical and clinical divide—identification of problems and overcoming obstacles[J]. Translational Medicine Communications, 2019, 4(1): 1-19.

统计显示 I 期到 II 期的平均转化率约为 52.0%。在 II 期阶段，10 个样本完成试验、4 个样本成功，成功率为 40%，优于行业平均水平（约 28.9%）。需要强调的是，上述统计基于公开披露且样本量较小，可能存在披露/幸存者偏倚，结论应谨慎外推<sup>95</sup>。典型案例是英矽智能（Insilico Medicine），其利用生成式 AI 平台识别出特发性肺纤维化（IPF）的新靶点（TNIK）并设计候选分子 Rentosertib（ISM001-055，原名 INS018\_055），实现了“18 个月内提名临床前候选化合物、30 个月内进入临床 I 期”的研发速度，显著优于传统模式下的研发周期。其 IIa 期临床试验结果显示，在安全性可接受的前提下，治疗组在第 12 周的用力肺活量（FVC）平均变化与安慰剂组呈方向性差异，提示探索性疗效信号，但仍需更大样本与更长随访进一步验证<sup>96</sup>。



来源：How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons

<sup>95</sup> Jayatunga M K P, Ayers M, Bruens L, et al. How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons[J]. Drug discovery today, 2024, 29(6): 104009.

<sup>96</sup> Xu Z, Ren F, Wang P, et al. A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: a randomized phase 2a trial[J]. Nature Medicine, 2025: 1-9.

图 8 AI 发现药物分子在临床试验中的成功率示意图

**二是中端闭环：在科技生物环节，AI 通过“高通量湿实验+自动化设计-合成-测试循环”强化数据生成与验证闭环，推动药物发现从“单点算法”走向“平台化系统能力”。**标志性事件是递归制药(Recursion)与 Exscientia 于 2024 年 11 月完成合并<sup>97</sup>，这一整合将 Recursion 大规模的“生物学表型探索能力”（每周超 200 万次湿实验）与 Exscientia 精密的“化学设计与自动化合成能力”相结合，构建生物学与化学端到端药物发现平台。在亚洲地区，晶泰科技(XtalPi)亦是该范式的典型代表，其于 2024 年在港交所上市，核心在于构建了大规模的自动化机器人实验室，通过“AI 预测+机器人验证”的干湿闭环加速药物固态研究与分子发现，并已获得礼来(Eli Lilly)等跨国药企的平台级合作订单。

**三是后端降本：在临床开发与决策环节，AI 通过“生成式模型驱动的软件化工程+临床运营优化”实现降本增效。**为缓解临床阶段的高昂成本，大型药企开始引入生成式 AI 重塑开发流程。赛诺菲(Sanofi)于 2024 年 5 月宣布与 OpenAI 及 Formation Bio 达成合作，目标是开发面向药物开发的 AI 软件与能力，以支持更广泛的研发流程与运营环节。随后，Formation Bio 推出了 AI 驱动的患者招募工具 Muse，体现了从“合作框架”走向“工具化落地”的路径。可量化证据层面，TrialGPT 等研究给出了在其研究设定下的效率收益(例如筛查时间显

<sup>97</sup> RECURSION PHARMACEUTICALS. Recursion Pharmaceuticals, Inc. completed the acquisition of Exscientia plc[EB/OL]. (2024-11-20)[2025-01-10]. <https://ir.recursion.com/news-releases/news-release-details/recursion-completes-acquisition-exscientia-creating-global>.

著减少)的实验结果,可作为“大模型用于临床试验匹配/预筛”的实证参考,但仍需结合真实世界流程与数据条件评估可迁移性<sup>98</sup>。

## 2.材料研发

新材料是科技和产业创新的基石,但**当前材料研发整体呈现“高投入、长周期、高不确定性”**的特征。权威评估普遍指出,一种新材料从概念设计、实验室合成、性能表征、中试放大到最终实现商业化应用,典型周期往往长达 10–20 年<sup>99</sup>,这一过程中需要长期投入昂贵的实验设备、试验材料和专业人才,前期资金沉淀大、回报节奏慢,且失败率较高。这导致了企业在既有材料体系上做渐进改良的结构性保守倾向,而不愿轻易押注颠覆性材料路线。这其中存在如下关键卡点,**一是多目标耦合与工程放大的挑战并存**。面向产业应用的材料研发,并非追求单一性能最优,而是要在性能、成本、可制造性和环境约束(如碳足迹、关键原材料依赖)之间进行高维多目标优化。很多在实验室小样条件下表现优异的材料,一旦走向放大生产,就会因设备差异、批次波动、工艺窗口收缩等因素导致微观组织和缺陷分布发生变化,引发性能不稳定甚至失效,导致企业需经历漫长的工艺迭代寻找折中方案。这种“实验室性能—工程可靠性”落差在电池、航空等安全敏感领域的影响尤为突出。**二是数据-模型-工艺割裂,数字化与标准化基础薄弱**。数据碎片化问题突出,实验、计算及工艺数据广泛散落在学术文献、纸质记录和企业各类孤立系统中,格式不统一、

<sup>98</sup> Jin Q, Wang Z, Floudas C S, et al. Matching patients to clinical trials with large language models[J]. Nature communications, 2024, 15(1): 9074.

<sup>99</sup> ROBINSON D K R, NADAL D. Steering the future of advanced materials: strategic intelligence in action[R/OL]. Paris: OECD, 2025-10[2025-12-12]. Available from: [https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/10/steering-the-future-of-advanced-materials\\_9add3995/a15874fa-en.pdf](https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/10/steering-the-future-of-advanced-materials_9add3995/a15874fa-en.pdf)

难以打通。尽管 Materials Project、NOMAD 等数据库已初具规模，但行业整体仍缺乏类似生物学 PDB( Protein Data Bank, 蛋白质资料库 ) 那样覆盖广泛、标准统一的“单一权威数据枢纽”。同时，企业研发流程的“手工化”倾向明显，大量研发活动仍高度依赖个人经验与试错，高价值数据资产难以有效沉淀与复用，难以与仿真模型、AI 工具及自动化装备形成闭环联动，成为制约材料产业实现数据驱动和智能化跃迁的关键瓶颈。

近年来，AI 赋能材料研发领域的进展主要有如下方面。首先是虚拟筛选与性质预测。以图神经网络、材料信息学 ( MI ) 为核心，AI 在庞大化学空间中进行高通量虚拟筛选，为电池、催化等产业提供“候选库”。DeepMind 的 GNoME 项目 ( 2023 年 ) 利用图网络和主动学习，对无机晶体进行高通量稳定性预测，提出约 220 万个新晶体结构，其中 38 万个被评估为计算热力学稳定。这类工作一方面显著扩展了候选空间，另一方面也把瓶颈更集中地推向“如何以更高吞吐完成合成与验证”。此外，微软与美国能源部西北太平洋国家实验室合作 ( 2024 年 ) 提供了一个“破局”示范，双方使用 AI 和 HPC 从约 3200 万种候选材料中筛选出一种新型锂钠固态电解质 ( 代号 N2116 )，该材料已被合成并在原型电池中验证，通过引入钠离子，可在保持导电性能前提下减少约 70% 的锂用量。这是少数已走完“AI 筛选 → 实验合成 → 原型电池验证”全流程的示范性案例之一，但距离大规模商业化仍处在较早的验证阶段。

其次是自驱动实验与闭环研发。将 AI 与自动化平台结合，形成

“AI 决策+机器人执行”的闭环，是“验证瓶颈”的关键。最具影响力里程碑是劳伦斯伯克利国家实验室研发的 A-Lab（2023 年），该自动化实验室在无人干预下，17 天内成功合成了 GNoME 预测的 41 种新材料，成功率达 71%，有力证明了“计算+自动化”闭环的有效性。在商业化层面，IBM（RoboRXN）、Emerald Cloud Lab、深云智合等企业，正在探索以云服务的形式将高通量自动化实验能力向外开放，为材料研发提供按需验证能力，并有望显著缩短验证周期。

**最后是逆向设计。**当前大规模工业化落地的应用仍多处于目标导向的筛选阶段，即“材料数据平台+反向查找+自驱实验”，其本质仍未脱离“前向预测+虚拟筛选”的技术路径。而使用 AI 的原生逆向设计——直接给出目标约束生成全新材料结构，再物理验证——尚处于研究探索阶段，例如微软提出的 MatterGen（2025 年）是目前最具代表性的生成式材料逆向设计模型之一，通过扩散模型可在全元素周期表范围内生成稳定无机晶体，但其当前仍主要停留在“结构-性质”层级，对可合成性、工艺放大与成本等工程约束尚未实现端到端建模。

### 3. 先进制造与工程设计

先进制造与工程设计广泛覆盖航空航天、汽车、能源装备、高端机械等领域，是连接科学突破与产业落地的关键环节，其研发（设计、仿真、工艺开发与试验设计）本身正面临一系列内生瓶颈。**一是经验驱动与黑盒工艺导致的“试错依赖”。**在工艺开发等环节，“实验—试错—微调”的路径依然普遍存在。复杂拓扑结构、新型散热方案等，需要大量试验验证；焊接、增材制造、热处理、成形等工艺中，参数

与性能之间存在高度非线性“黑盒”关系，现实中高度依赖少数资深工程师的隐性知识。相关规律难以系统建模、难以在团队间传承，造成开发周期长、可复制性弱。**二是极端复杂性与高昂仿真成本导致“算得太慢”。**现代工程系统往往是强耦合、多物理场、多尺度的复杂系统，高保真计算流体力学( Computational Fluid Dynamics, CFD )仿真、有限元分析 ( Finite Element Analysis, FEA ) 以及热-流-固耦合 ( Thermo-Fluid-Structure Interaction, TFSI ) 仿真是设计评估的必选项，但单次仿真往往需要数十甚至数百小时，导致设计空间难以被充分探索。工程师只能在有限工况和参数组合上做“点状”试探，大量决策仍依赖经验与安全裕度，难以实现真正的全局最优。**三是串行流程与“设计 - 制造”鸿沟导致的“协同不畅”。**传统研发流程高度串行，先在计算机辅助设计 ( Computer-Aided Design, CAD ) 环境中完成结构与外形设计，再交由计算机辅助工程 ( Computer-Aided Engineering, CAE ) 团队做仿真验证，最后才由工艺/制造团队评估可制造性与成本。一旦在仿真或试制阶段发现结构薄弱、工艺不可行或成本过高，只能回滚前端设计，修改成本极高，周期被大幅拉长。可制造性、供应链可获得性、全生命周期成本与碳排约束普遍被“后置”，出现“纸面设计优秀但造不出来、造出来不经济”的现象，也难以支撑小批量、多品种、定制化的市场需求。

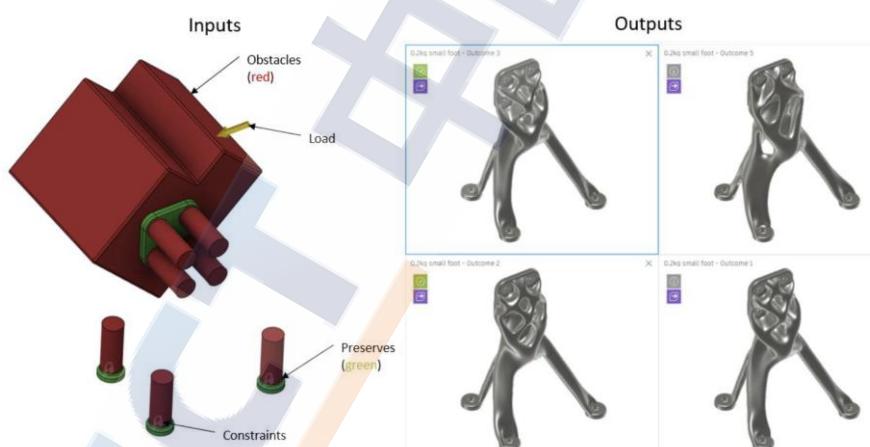
近年来 AI 在工程研发中的应用大致可以归纳为以下方向。首先**是约束感知生成式设计**。工程师不再手工绘制结构，而是输入载荷、边界、材料、成本以及制造方式等约束条件，AI 在给定设计空间中自

动生成大量候选方案，再通过仿真与工程审查筛选出可行解。最新的趋势是“约束感知”：在生成阶段就同时考虑制造可行性（如特定 3D 打印工艺的限制）、供应链可获得性以及全生命周期成本和碳排，使得生成结果不再停留在“理论上最优”，而是可以直接进入工程化落地。美国国家航空航天局戈达德航天中心（NASA Goddard）自 2022 年起提出进化结构（Evolved Structures）流程，通过将生成式设计、有限元分析和数字制造一体化，在 EXCITE 气球望远镜 Tip/Tilt 支架的研发中，AI 生成结构在质量相近的前提下，不仅性能更优，刚度/质量比提升超 3 倍，最大等效应力降低约 7–9 倍，设计效率提升一个数量级，1 名工程师约 1.5 小时完成 2 名工程师 2 天的工作量。该技术已在光学平台 ALICE、X 射线探测器 STAR-X 以及火星样本返回等任务的关键支架和仪器承力件上应用<sup>100</sup>。美国 Divergent 公司研发的自适应生产系统（Divergent Adaptive Production System, DAPS）以拓扑优化、生成式设计和物理仿真为内核，自动生成满足强度、刚度、碰撞、安全等多目标约束的结构拓扑；再通过金属增材制造一体成形关键节点结构，最后由机器人完成装配，实现“设计—仿真—制造—装配”的一体化流程。在汽车领域，DAPS 以 Czinger 21C 为首个整车级应用示范，21C 的底盘与关键车身结构由数百个金属打印节点和杆件构成，通过算法生成的仿生拓扑在满足赛道级强度与刚度的同时，实现极致轻量化<sup>101</sup>。2024 年布加迪利用 DAPS 开发新车型陀飞轮

<sup>100</sup> McClelland R. Generative design and digital manufacturing: using AI and robots to build lightweight instrument structures[C]//Current Developments in Lens Design and Optical Engineering XXIII. SPIE, 2022, 12217: 141-148.

<sup>101</sup> Schuhmann D, Heß F, Heß S, et al. A study on additive manufacturing for electromobility[J]. World Electric Vehicle Journal, 2022, 13(8): 154.

(Tourbillon)，将多连杆前后悬架的控制臂和转向节替换为 AI 设计的有机形态铝合金 3D 打印构件，相比前代车型悬挂总成减重约 45%，并在后悬架上采用 AI 设计的中空机翼型摆臂，在提升空气动力学性能的同时进一步降低非簧载质量<sup>102</sup>。在航空与防务领域，Divergent 与通用原子航空系统公司 (GA-ASI) 合作，用 DAPS 为一型小型无人机设计和制造整体机体结构，将原本约 180 个分离零件集成为 4 个复杂 3D 打印节点，零件数量减少超过 95%，单个节点打印时间控制在 13 小时以内，四节点机体的机器人总装时间不足 20 分钟，该类项目在开发成本上可节省约 50%，经常性制造成本则有望降低 55%–75%<sup>103</sup>。上述案例也标志着生成式设计和增材制造正在迈向工程主流路径。



来源：Generative design and digital manufacturing: using AI and robots to build lightweight instrument structures

图 9 生成式设计流程示意图（左侧输入设计目标，右侧生成方案）

<sup>102</sup> Light Metal Age. Bugatti Tourbillon features aluminum components for improved performance[EB/OL]. 2024-07-02[2025-12-07]. Available from: <https://www.lightmetalage.com/news/industry-news/automotive/bugatti-tourbillon-features-aluminum-components-for-improved-performance/>

<sup>103</sup> Scott C. Additive manufacturing and artificial intelligence combine for an automated solution[EB/OL]. Wohlers Associates, 2024-09-03[2025-12-07]. Available from: <https://wohlersassociates.com/uncategorized/additive-manufacturing-artificial-intelligence-combine-automated-solution/>

**其次是 AI 代理模型与加速仿真。**以物理信息神经网络、图神经网络、神经算子为代表的新一代代理模型，利用高保真 CAE 仿真或实验结果为训练数据，学习系统的物理规律，在可控误差范围内实现数十倍乃至数百倍的加速。在此基础上，研发流程可以从“设计→漫长仿真→修改”的串行，变为“实时设计↔实时近似仿真”的高频交互。例如，英伟达推出代理模型训练框架 PhysicsNeMo<sup>104</sup>，澳汰尔（Altair）、安西斯（Ansys）等 CAE 厂商快速将其集成到仿真产品。在复杂外流场场景中，Altair 研发的 AI 代理模型可以将单次高精度 CFD 仿真从约 750 分钟压缩到 3 分钟左右，且误差控制可接受<sup>105</sup>。云仿真平台 SimScale 推出的离心泵模型可在 1 秒内。预测泵效率和扬程，相比传统 CFD 求解实现约 2700 倍加速，实现 AI 代理模型与云仿真平台的闭环验证<sup>106</sup>。与此同时，Altair、Ansys、西门子等主流 CAE 厂商纷纷推出面向工程师的 AI 副驾驶功能，通过自然语言理解需求、自动推荐模型设置与网格策略、辅助后处理与结果解释，降低仿真使用门槛，减少重复性操作，帮助用户聚焦高价值模型构建。

**第三是数字孪生与系统级研发决策。**数字孪生技术正从设备级监测走向系统级研发决策平台。在整车工厂、造船厂、复杂能源系统等重大工程项目中，数字孪生与 AI 结合，用于评估不同布局方案、工艺路线与投资组合，帮助决策者在投入巨大资本支出前完成系统性推

<sup>104</sup> NVIDIA. NVIDIA PhysicsNeMo[EB/OL]. [2025-12-05]. <https://docs.nvidia.com/physicsnemo/>

<sup>105</sup> Altair Engineering Inc. Applications of ML/AI for CAE[R/OL]. 2024-02[2025-12-05]. [https://revolutioninsimulation.org/wp-content/uploads/2024/02/Altair\\_NAFEMS\\_Feb2024\\_v3.pdf](https://revolutioninsimulation.org/wp-content/uploads/2024/02/Altair_NAFEMS_Feb2024_v3.pdf)

<sup>106</sup> SimScale. SimScale Unveils the World's First Foundation AI Model for Centrifugal Pump Simulation Built with NVIDIA PhysicsNeMo[EB/OL]. 2025-03-18[2025-12-05]. Available from: <https://www.simscale.com/press/simscale-unveils-worlds-first-foundation-ai-model-centrifugal-pump-simulation-built-nvidia-physicsnemo/>

演。宝马联合英伟达研发了数字孪生平台 FactoryExplorer，将建筑、设备、物流、车型配置和人工操作等数据统一到高保真三维数字孪生中，并在其中集成自动碰撞检查、人因工效仿真和智能物流等 AI 能力。到 2025 年，该平台已在宝马全球 30 余座工厂部署数字孪生，计划在 2027 年前为 40 余款新/改款车型完成虚拟导入，关键的新车型碰撞检查从过去近 4 周压缩到约 3 天完成，有望将生产规划成本最高降低约 30%<sup>107</sup>。

#### 4. 半导体与芯片设计

半导体是数字经济的基石。然而，随着工艺节点推进至 3 纳米及以下，“摩尔定律”遭遇物理、经济与系统复杂性的多重极限，研发挑战日益严峻。**一是物理极限与经济极限叠加。**芯片制造面临“功耗墙”与“成本墙”的双重制约。晶体管尺寸逼近物理极限，量子隧穿效应与漏电流导致功耗和散热压力剧增，单纯缩小尺寸不再自动带来性能、功耗、面积指标的线性提升。同时，先进制造工厂投资动辄百亿美元级别，极紫外光刻等尖端设备、掩膜版制作与计算光刻成本高昂，使得单次流片失败的代价可达数千万美元，研发试错空间被极度压缩。**二是设计与验证复杂度爆炸。**先进系统级芯片集成数百亿晶体管，包含多种处理核心与复杂接口，在设计空间呈指数级膨胀的背景下，性能、功耗、面积、时序等多目标优化已远超人工经验调参的能力范围。电子设计自动化工具（Electronic Design Automation, EDA）

<sup>107</sup> BMW Group. BMW Group scales Virtual Factory[EB/OL]. (2025-06-11)[2025-12-06]. Available from: <https://www.press.bmwgroup.com/global/article/detail/T0450699EN/bmw-group-scales-virtual-factory?language=en>

所处理的逻辑综合、布局布线等问题本质上是难解的组合优化问题，传统算法难以兼顾优化质量与收敛速度。更为突出的是，验证环节消耗了芯片项目大部分的人力和时间成本，其状态空间的急剧扩张使得彻底排查设计缺陷近乎不可能，验证不足导致的设计返工与潜在风险显著上升。**三是工艺与设计工具不确定性带来的技术约束。**先进工艺本身的不确定性增加，到了 3 纳米等节点，工艺波动、器件差异和版图依赖效应更明显，仿真难以覆盖所有情况，设计团队只能被动增加时序、可靠和可制造性裕量，导致性能与能耗优化空间被不断压缩。现有设计工具和模型存在能力边界，传统电子设计自动化在多物理场耦合、统计时序分析、变异感知优化等方面仍有精度和计算瓶颈，无法对所有“角落场景”做充分分析。工程师不得不在分析精度、计算成本和项目周期之间反复权衡，架构和版图决策的不确定性随之提高。

AI 赋能芯片研发与设计正在从“局部尝试”走向“主流程重构”，主要体现在物理实现、验证模拟及架构探索等方向。**首先是辅助实现与物理设计，这是 AI 在 EDA 领域落地最成熟、商业价值最突出的环节，聚焦于从寄存器传输级（Register-Transfer Level, RTL）到物理版图阶段的功耗、性能与面积（Power, Performance, Area, PPA）自动优化。**新思科技（Synopsys）推出的 DSO.ai 工具使用强化学习在既定工艺下自动探索综合、布局布线组合。自 2020 年推出以来，其被广泛采用，已用于超 300 次商业流片。在不同的先进节点典型项目中，可实现功耗降低约 25%；最高主频提升 4.5%；芯片面积缩小 6.5%；设计探索效率提升约 3 倍以上，相当于将资深工程师多年的设计调优经

验封装成了可复用的智能系统<sup>108</sup>。楷登电子( Cadence )推出的 Cerebrus 平台同样基于强化学习优化全流程，已有超过 1000 家客户使用该平台完成了 28 纳米及以下工艺的流片<sup>109</sup>。据联发科披露，通过 Cerebrus 优化 SoC 设计，在典型项目中总功耗可下降 12.5%，芯片面积缩小约 5%、整体优化周期缩短 50%<sup>110</sup>。上述案例说明，AI 驱动的物理设计优化已从可选项变成主流生产力。同时，谷歌 AlphaChip 将平面布局建模为强化学习过程，2020 年发布以来，AlphaChip 已应用于谷歌多代张量处理单元 ( Tensor Processing Unit, TPU ) 和数据中心 CPU 的设计，能在数小时内生成媲美甚至超越工程师多周成果的布局方案，在连线长度可降低约 5%<sup>111</sup>，这被视为 AI 直接承担关键物理设计任务的标志。

**第二是辅助验证与测试，AI 在此环节主要用于回归与覆盖优化、调试与根因分析。**新思科技推出的验证工具 VSO.ai，通过机器学习分析覆盖率数据，自动发现冗余并建议高价值用例，可加快覆盖收敛、减少测试次数与仿真资源消耗。根据英伟达在 3 个芯片项目中的应用分析，验证功能覆盖率最高提升 33%，测试平台回归压缩率达 2-7 倍，且可发现独特缺陷<sup>112</sup>。楷登电子也推出类似工具 Xcelium ML，在不

<sup>108</sup> SYNOPSYS, INC. Form 8-K: Current Report [R/OL]. Washington, D.C.: U.S. Securities and Exchange Commission, 2024-03-20 [2025-12-08]. <https://www.sec.gov/Archives/edgar/data/883241/000119312524072741/d785364d425.htm>

<sup>109</sup> CADENCE. Chip Design Industry Reaches an AI Inflection Point [EB/OL]. (2025-05-08) [2025-12-09]. [https://community.cadence.com/cadence\\_blogs\\_8/b/corporate-news/posts/chip-design-industry-reaches-an-ai-inflection-point](https://community.cadence.com/cadence_blogs_8/b/corporate-news/posts/chip-design-industry-reaches-an-ai-inflection-point).

<sup>110</sup> KHERA V. What's Behind the 5% Die-Area Shrink and 12% Power Saving by MediaTek? [EB/OL]. (2022-08-22) [2025-12-09]. [https://community.cadence.com/cadence\\_blogs\\_8/b/di/posts/cerebrus-with-ai--improves-ppa-and-productivity-for-mediatek](https://community.cadence.com/cadence_blogs_8/b/di/posts/cerebrus-with-ai--improves-ppa-and-productivity-for-mediatek).

<sup>111</sup> Goldie A, Mirhoseini A. How AlphaChip transformed computer chip design [EB/OL]. 2024-09-26 [2025-12-09]. Available at: <https://deepmind.google/blog/how-alphachip-transformed-computer-chip-design/>

<sup>112</sup> Reddy T. How NVIDIA uses functional verification tools VSO.ai to accelerate coverage closure [EB/OL]. 2024-09-04 [2025-12-09]. Available at: <https://www.synopsys.com/blogs/chip-design/vso-ai-nvidia.html>

牺牲覆盖率的前提下，瑞萨电子使用该工具将回归测试集压缩到原来的约一半，并把功能覆盖率恢复到 100%；在后续派生版本上，测试规模进一步压缩到原来的约 25%，同样实现了 100% 覆盖，从而在保证验证质量的同时，降低验证成本与周期<sup>113</sup>。同时，业界也推出了调试与根因分析工具，使用 AI 分析失败用例以定位高价值问题，如新思科技的 Verdi 系统和楷登电子的 Verisium 平台，可以自动完成“失败分组—嫌疑范围收窄—根因路径挖掘”的闭环，使研发人员专注缺陷修复与设计优化。

**第三是架构探索与设计副驾**，在中游物理实现和下游验证测试之外，AI 也开始探索进入前端工作。一条路径是用 AI 做架构与微架构设计空间探索，自动搜索核心数、缓存、互连等参数，并用性能/功耗预测模型快速评估，缩小搜索空间，让架构师聚焦少数高潜力方案。另一条路径是设计副驾（design copilot），将大模型嵌入 EDA 平台，支持自然语言配置约束和脚本、辅助代码与规范审查、生成报告草稿，显著减少查资料和调试时间。整体来看，上述工作已经成为迈向设计智能体的过渡形态。

## 六、发展挑战与展望

### （一）发展挑战

科研智能作为人工智能在科学研究与技术研发中的深度应用，正在从工具升级走向范式重塑。然而，一个以数据驱动、工程导向、快

<sup>113</sup> Cadence Design Systems, Inc. Renesas and Cadence: boosting productivity by 6X by leveraging Xcelium ML and Verisium platforms [EB/OL]. 2023-07 [2025-12-09]. Available at: [https://www.cadence.com/en\\_US/home/resources/success-stories/renesas-and-cadence-ss.html](https://www.cadence.com/en_US/home/resources/success-stories/renesas-and-cadence-ss.html)

速迭代为特征的 AI 范式，在融入强调机理驱动、严谨求证、渐进累积的传统科研体系之中时，会不可避免地产生系统性摩擦。尤其是在从试点示范迈向规模化产业化的过程中，这些矛盾被进一步放大，成为制约科研智能发展的关键卡点。总体来看，当前挑战可归纳为数据、算法、工程、组织和治理五个维度。

**一、数据与知识基础薄弱，高质量“科学数据”与可嵌入知识仍显不足。**供给侧方面，与互联网场景相比，科学与研发领域的数据获取成本高、周期长，如高能物理、新材料、生物医药等领域的数据生产依赖昂贵的大型实验装置和漫长的实验周期，导致数据稀缺。数据孤岛和数据沼泽现象普遍，因知识产权、商业机密和体制障碍，数据长期分散在各实验室、科研机构和企业内部，FAIR 原则在一一线落地难。应用侧方面，存量数据普遍存在噪声大、偏差多、标注缺失、失败数据缺失等问题，且格式标准各异，缺乏统一规范。多模态科研数据（结构、光谱、图像、日志、文本等）的统一表征和建模仍处于探索阶段。同时，大量领域知识仍以教材、论文、实验记录和专家经验等非结构化形态存在，可复用的本体、知识图谱和领域知识库建设滞后。结果是，科研智能在进入新学科、新行业或新场景时往往需要从头做数据治理与知识梳理，工程成本高、可迁移性差，也难以沉淀成可定价、可交易、可持续运营的数据与知识资产。

**二、算法与可靠性不足，机理与智能融合尚未跨越。**现有主流模型本质上仍是统计相关性工具，对守恒定律、对称性约束、反应机理、失效机理等领域知识缺乏显式表达，外推能力弱，一旦超出训练分布

或进入极端工况，容易给出违背物理规律或工程常识的结果。深度模型“黑箱”属性突出，难以回答“为什么”“什么条件会失效”，与科研强调的因果解释、机理可验证和路径可追溯存在根本张力。大模型在文献综述、科研问答等场景中仍普遍存在幻觉，会捏造文献、数据和结论；AI 社区自身也存在代码不开源、数据泄漏、结果难复现等问题。将这套实践直接迁入科学研发和产业研发，容易放大伪发现、虚高结果和可重复性危机。在药物研发、关键材料、工程安全等高风险、高监管领域，如果缺乏可解释性、不确定性刻画和严谨评估体系，企业和监管部门都难以真正把科研智能纳入关键决策链条，相关系统只能停留在试用和辅助参考的边缘位置。

**三、工程化与产业化基础薄弱，从原型到产品的“最后一公里”尚未打通。**目前科研智能工具链高度碎片化，严重依赖各类脚本和零散工具，缺乏面向科学与产业研发场景的一站式平台和工作流体系。科研智能系统需要与高性能计算平台、学科软件、实验仪器、实验信息管理系统、企业研发管理和生产管理系统等既有基础设施深度耦合，接口多、改造重、复用性弱。算力与平台鸿沟也较为突出，中小科研机构和多数企业研发部门缺乏大模型训练、复杂仿真与自动化实验所需的稳定算力与专业运维能力。产业化层面，学术原型到工业级产品之间存在明显“死亡之谷”：要满足 $7 \times 24$ 小时稳定运行、合规审计和行业标准，需要大量额外工程投入，而现有项目制、短周期资助难以支撑。尽管中国人工智能产业发展联盟科学智能工作组、工业和信息化部人工智能标准化技术委员会已启动科研智能标准化研究，行业

仍然缺乏系统化的评估与采购标准，企业难以客观比较不同方案的实际价值，商业模式仍以定制化项目服务为主，可复用平台产品有限，产业链上下游接口和分工模式尚未成型。这些因素共同导致科研智能容易停留在演示系统和示范工程阶段，难以沉淀为可复制、可运营、可持续演进的“新型科研基础设施”。

**四、组织、人才与体制匹配不足，范式变革的“慢变量”制约显现。**科研智能本质上是高度交叉、强工程特征的系统工程，需要大量既懂AI又懂学科、既懂科研又懂工程和应用场景的“AI+X”复合型人才。但现实中，高校专业划分与培养路径仍沿传统学科线展开，AI人缺乏对领域问题的深度理解，领域科学家与工程师对AI方法、数据治理与算力平台又相对陌生，真正能统筹问题定义、数据工程、模型研发和工程落地的“双栖”主导型人才十分稀缺。组织结构上，“课题组+学科院系”的模式更适合小团队、单学科课题，对建设跨学科平台、跨机构联合攻关以及长期运营科研智能基础设施的支持不足。评价和激励体系仍围绕论文和传统成果展开，高质量数据集、软件平台、工具链和标准制定等基础工作在职称、项目和奖励中权重偏低，难以吸引和留住工程化与平台化人才。认知和文化层面，一端是对AI的怀疑与保守，一端是技术乐观主义和盲目跟风，都会削弱机构对科研智能的理性布局和持续投入，使这一新范式长期处于“重要但不紧迫”“想做但做不深”的状态。

**五、科研伦理与安全治理框架尚未成熟，“向善可控”的边界有待厘清。**需要重点关注的是“双重用途”风险，科研模型既可用于药

物、材料和工艺创新，也可能被滥用于设计危险分子、敏感材料或高风险工艺。生成式 AI 可能降低虚假内容与学术不端的实施门槛，伪造文献、捏造数据和伪发现风险上升，对科研共同体的可复现性与学术诚信提出更高要求。此外，科研模型可能存在偏见与歧视，AI 生成成果的知识产权面临归属难题，AI 在自动化实验室中的幻觉可能引发真实世界安全事故，这些都对现有治理框架提出严峻考验。目前，针对科研智能的专门伦理规范、访问控制、审计与追责机制仍不健全，跨国、跨机构的协同治理亦处在起步阶段。若相关风险缺乏有效治理，可能在舆论与监管层面引发对技术应用的审慎乃至收紧，从而影响科研智能的推广速度与应用范围。

## （二）未来展望

科研智能作为新一轮科技革命和产业变革的重要引擎，其意义已远远超出“提高科研效率的辅助工具”，正在成长为重塑知识生产方式和研发组织体系的基础性力量。基于本报告前序章节的分析，可以预期，在未来相当长的一段时期内，科研智能将沿着五条相互交织、彼此强化演进主线加速发展，共同描绘科学发现与产业研发的新图景。

**一、科研范式从人机协同加速迈向自主科研。**科研智能的深度应用首先体现在科研范式的跃迁上：从以人类研究人员为中心、AI 辅助的“人机协同”，逐步走向以智能系统为内核、可端到端执行的“半自主乃至自主科研”。在当前阶段，AI 主要扮演“智能副驾驶”“能力放大器”的角色，在研究人员设定目标和方法的前提下，高效承担文献检索、数据处理、模拟计算等环节性工作。面向未来，演进方向

是构建能够完成完整科研闭环的自主科研系统：由一个或多个高级科研智能体组成，在明确边界和安全约束下，自动完成“提出可检验假说—设计并优化实验方案—调度自动化平台执行实验或仿真—分析结果并更新模型—迭代修正认知”的端到端流程。这也将重构科研分工，人类研究员会更聚焦于提出高价值问题、审核关键结论和综合跨学科证据，智能系统承担大规模试错、复杂流程编排和持续优化，在整体上推动科研活动的速度和质量实现显著提升。

**二、模型体系走向大小模型协同与多模态机理深度融合。**一方面，大小模型协同将成为主流形态：大型基础模型作为通用知识平台和推理引擎，在多学科、多模态数据上预训练，提供跨领域知识整合与复杂推理能力；小型专家模型则围绕特定学科、特定任务或特定设备，基于高质量垂直数据进行精调，追求更高精度、更低成本和更强可解释性，实现大模型赋能、小模型落地的良性生态。另一方面，模型将加速走向多模态与机理融合：未来科研大模型必须能够同时处理文本、图像、序列、结构等多源异构信息，并通过在模型结构、训练目标和损失函数中显式融入物理定律、化学规则、工程约束和因果结构，将数据驱动与机理推理有机结合。只有在多模态理解与科学机理深度嵌入的前提下，系统的输出才有可能真正满足研究人员对严谨性、外推性和可解释性的高要求，在应用场景中获得长期信任。

**三、科研基础设施设施平台化，演进为“科研资源即服务”（Research as a Service, RaaS）。**科研大模型和自动化实验室的高昂成本，正在推动科研基础设施供给模式发生根本性变革，从“各自建

设、自行使用”的分散格局，转向以平台为载体的服务化体系。“科研资源即服务”模式的核心，是通过云平台，将分散的算力、数据、模型和实验设施抽象为可组合的服务能力，以按需、弹性、标准化的方式向科研机构和企业开放。面向用户，这一服务目录将包括：面向科研工作负载优化的“算力即服务”，符合 FAIR 原则的“数据即服务”，可调用与微调的“模型即服务”（Model as a Service, MaaS），以及可远程访问和编程控制的“实验即服务”（Experimentation as a Service, EaaS）。RaaS 模式使得资金有限的中小型研究团队、初创公司乃至个人研究者，也能以较低的门槛获取顶级的科研能力，从而极大地激发全社会的创新活力。同时，其有望催生一批围绕平台运营和生态建设的新型科研基础设施服务商，成为国家创新体系和产业体系的重要底座。

**四、产业化与场景规模化推动千亿级应用赛道系统崛起。**科研智能的潜力和价值需要在实际研发场景中的规模化应用来体现，未来将从当前的单点突破（如 AlphaFold），走向贯穿产业研发全链条的系统性落地，在若干关键行业形成千亿级甚至更大规模的新兴赛道。一方面，AI 将从单一环节工具，演化为覆盖需求洞察、机理建模、方案设计、工艺优化、中试放大到验证测试的端到端研发平台，深度嵌入企业新产品、新材料、新工艺的形成过程。另一方面，在药物、新材料、新能源、半导体与先进制造等领域，“AI+自动化实验”的闭环模式有望成为研发新常态：药物研发有望实质性改写传统“双十定律”，材料反向设计将服务于电池、光伏、催化与高端合金等战略方向，AI

将成为 EDA、CAD/CAE 等工业软件的核心算法引擎，推动复杂产品设计、验证与优化走向高度自动化。围绕这些场景，将逐渐形成算力与基础模型供应商、行业科研平台提供商、应用集成商与专业服务商的完整产业链，涌现一批具有全球竞争力的科研智能平台型企业。

**五、可信与治理体系致力于构建负责任的创新生态。**随着 AI 在重大科研项目和高风险工程研发中的分量不断上升，其可信性、安全性和伦理性不再是外围议题，而是关乎能否在关键领域被采纳和长期使用的基础条件。在技术层面，需要持续推进可解释性 AI、不确定量化和鲁棒性评估等方法，确保科研模型的推理路径可理解、预测结果有置信度界定、输出行为符合基本的物理机理和工程常识。在治理层面，需要面向科研活动建立系统的规则与机制，例如数据治理标准与访问控制、算法审计与可复现机制、AI 参与科研过程的记录与证据链规范，以及围绕 AI 生成内容和 AI 参与发现的知识产权规则等。同时，针对 AI 在生物、化学等领域的双重用途风险，需要通过科技伦理审查、安全评估和跨国协同治理等方式建立坚固护栏，确保科研智能在推动原始创新与产业升级的同时，始终在安全可控、负责任的轨道上运行。

## 编制说明

本研究报告自 2025 年 9 月正式启动编制，分为前期研究、文稿起草、征求意见和修改完善五个阶段。面向科研智能领域的相关机构开展了深度访谈和调研等工作。

本报告由中国信息通信研究院人工智能研究所、中国人工智能产业发展联盟联合撰写，撰写过程中得到了华为技术有限公司、曙光信息产业股份有限公司、新华三技术有限公司、上海人工智能实验室、北京大学、中国科学院深圳先进技术研究院、北京低碳清洁能源研究院、之江实验室、北京百度网讯科技有限公司、中化信息技术有限公司、北京枫清科技有限公司等单位的大力支持。

中国信息通信研究院 人工智能研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62301618

传真：010-62301618

网址：[www.caict.ac.cn](http://www.caict.ac.cn)

