引用格式：洪学海. 构建AI原生创新体系，推动科研组织管理模式转型创新. 中国科学院院刊, 2026, 41(7): 1403-1415, doi: 10.3724/j.issn.1000-3045.20260303005.
Hong X H. Building AI-native innovation system to drive transformation and innovation in research organization and management models. Bulletin of Chinese Academy of Sciences, 2026, 41(7): 1403-1415, doi: 10.3724/j.issn.1000-3045.20260303005. (in Chinese)

构建AI原生创新体系，推动科研组织管理模式转型创新

洪学海

中国科学院计算技术研究所　北京　100190

摘要　人工智能（AI）正在深刻重塑科研范式。文章旨在解析AI赋能科研的内在机理及其对科研组织管理模式的影响。通过AI赋能科研的“能”与“不能”归纳，揭示了当前AI赋能科研的成效与能力边界。基于对AI赋能科研共性核心条件提炼和AI赋能的科研组织模式的典型案例解构，分析了AI赋能科研的组织管理模式与传统科研组织模式差异。同时提出了AI赋能科研需要破解有组织科研与自由探索、平台化供给与个性化需求、智能化科研生产力与传统科研管理制度三大矛盾。初步提出了构建AI原生科研创新体系，推动科研组织管理模式转型与创新的机制与路径。

关键词　人工智能，科研范式，科研组织管理，AI原生，创新体系

DOI　10.3724/j.issn.1000-3045.20260303005

CSTR　32128.14.CASbulletin.20260303005

人工智能（AI）赋能科研已经取得了一系列重大科技成果。例如，2020年Google DeepMind公司[①]（以下简称DeepMind）发布的蛋白质结构预测工具AlphaFold2，颠覆了人们对AI在科学领域所能达成成就的预期，被*Science*杂志评为2021年度科学突破，并荣获2024年诺贝尔化学奖。当前，AI正从科研“辅助工具”演变为“科研协作伙伴”，甚至演变成为自主型科学发现的系统“AI科学家”，使AI技术成为驱

资助项目：中国科学院战略研究与决策支持系统项目（GHJ-ZLZX-2026-10）
修改稿收到日期：2026年7月8日

① DeepMind由戴密斯·哈萨比斯(Demis Hassabis)等于2010年在英国伦敦创立，于2014年被谷歌收购，此后作为独立公司运营。2023年4月，谷歌母公司Alphabet宣布将Google Brain（于2011年成立）和DeepMind合并为Google DeepMind，以应对OpenAI与微软等在AI领域的竞争压力。

动科技创新的核心引擎之一。

然而，AI赋能科研不是简单“AI+科研”或“科研+AI”，更不是AI替代人的全部科研过程，而是将AI模型、算法及Agentic AI等直接熔合到具体科技领域科学技术问题求解的方案设计、假设验证、逻辑推演、结果分析等环节，突破了传统“+”的简单组合模式。随着AI算法不断创新、大模型迭代日益提升能力，以及Agentic AI深入渗透科研过程，AI赋能科研的潜力边界正在被不断拓展。这不仅是科研工具的革新，更是一场触及科研治理体系的深刻革命。未来，传统科研组织管理模式需及时进行适度改革，要构建AI原生创新体系，以适应AI带来的科研范式革命，从而提升科研组织的战斗力和创新体系整体效能，增强国家整体科技竞争力。

1 AI赋能科研的“能”与“不能”

1.1 AI能极大地提升科研效率

*科研效率是科研能力的重要表现之一。*2024年5月英国皇家学会发表的报告[1]，调查了AI如何改变科学研究，得出的主要结论是：AI应用可见于所有科学、技术、工程、数学（STEM）领域，并正在深刻重塑科学研究的范式、方法论和效率。Tshitoyan等[2]和Eger等[3]调查指出，AI能超越人类阅读极限，能实时监控、分析与综合全球每年数百万篇新发表论文，揭示跨学科、跨时间尺度的宏观研究趋势与“暗知识”，可以快速梳理特定领域的知识脉络、核心争议、研究空白、生成新假设和启发全新交叉研究课题等。2025年*Science*评选出年度十大科学突破，其中第7项是“大语言模型助力科学研究”，*Science*评论指出：大语言模型（LLM）已经在众多科学领域展现出了卓越的专业能力。例如，在数学领域，OpenAI公司的GPT-5在组合数论和图论中解决了困扰数学家数十年的难题；在化学领域，微调后的LLM仅用15次实验就为一种未知的复杂反应找到了最佳条件，省去了数百次耗时数周的实验。

1.2 AI能有效应对部分复杂科学问题

目前，AI赋能科研的内在机理模式有“数据驱动、知识驱动和数据与知识融合增强”3种，核心在于将模糊的科学问题转化为可计算的优化与预测任务。李国杰[4]将AI赋能科研领域的科技问题求解特点概括为6点，其中有2点指出了用AI求解的科技问题特征：① 以复杂系统为主要研究对象，有效应对计算复杂性非常高的组合爆炸问题；② 面向非确定性问题，概率和统计推理在科研中发挥更大的作用。孙凝晖[5]指出，AI在突破人类认知极限上，有无确定物理模型的数据驱动和有确定物理模型的机理内嵌2条典型技术路线。本文梳理了相关领域成果案例，如信息技术[6,7]、化学[8]、材料科学[9-12]、生命科学[13]、凝聚态物理[14,15]、天文学[16]和地球科学[17]，显示了AI赋能科研3种内在机理模式在应对复杂系统和非确定性问题方面具有显著潜力，能够通过数据驱动、物理机理内嵌等技术路径进行应用。这正是AI将数据驱动的“黑箱”探索能力，与机理内嵌的“白箱”约束能力相结合，以应对复杂系统的组合爆炸和非确定性问题带来的挑战。DeepMind基于Gemini大模型和Aletheia智能体（AI数学家，AI数学研究系统）连发多篇数学方面的论文[18-20]，包括Erdős猜想、Max-Cut（最大割）、Steiner Tree（斯坦纳树）等数学问题。其中关于Erdős猜想，涉及大量的图论与数论问题，这些问题涉及海量的可能性组合，计算复杂度极高，Max-Cut、Steiner Tree问题属于非确定性多项式时间难（NP-hard）问题，搜索空间极其庞大，具有典型的“组合爆炸”特征。AI数学家利用大模型的通用推理能力（知识驱动），进行符号推理和逻辑演绎来生成答案，证明了AI大模型开始具备在高度复杂的离散结构和符号系统中进行逻辑推理和搜索的能力。

1.3 AI赋能科研的能力边界

上述AI赋能科研取得的成就，可能并不意味着当前AI能够解决所有科学技术问题。目前AI能够解决科学问题的共性特征是：目标函数明确、解空间清晰可控、高质量数据支撑、计算可并行重复、科学规律具时空稳定性、结果评价指标可量化。事实上，AI的能力存在明确的边界，这些边界由计算复杂性理论、认知科学原理及物理世界的根本特性所决定。目前AI无法解决的科学问题的共性特征是：① 计算复杂性本质限制。虽然AI已能通过数据驱动和机理内嵌的技术路径，在芯片设计等NP-hard问题上找到超越人类经验的帕累托最优解，或在材料筛选中应对高维复杂空间，但这主要是在实践中采用近似求解、启发式搜索或数据驱动预测的方式来“有效应对”这些问题，而非在理论上攻克NP-hard的计算复杂度本质[21,22]。② 理论框架缺失。如量子引力、意识本质等问题面临理论框架的不完备。解决这些问题需要理论创新而非单纯依赖数据增加和算力提升。在缺乏物理模型或领域知识支撑的“非确定性”领域，AI难以独立完成理论突破[23,24]。③ 实验验证困难。某些问题涉及无法实施的实验或无法获取的数据。例如，普朗克尺度、早期宇宙、生命起源等问题的研究缺乏实验反馈，导致理论进展困难，AI也因缺乏有效输入而难以发挥作用[25,26]。④ 数据稀缺性障碍。如稀有事件、长尾分布、历史唯一性事件等问题面临数据缺失。当数据本质上不可获取且无法通过物理机理约束进行补偿时，AI方法往往失效[27,28]。从计算理论角度看，哥德尔不完备性定理对AI意味着：任何基于规则的AI系统都有其无法解决的问题，存在AI无法证明但人类可以“看出”的真理，数学“洞见”超越算法。图灵停机问题不可计算性也是对AI能力的制约。此外，认知神经基础尚不完全清楚热力学和量子力学等物理定律的限制。这些共同决定了当前AI赋能科研的能力边界。

2 AI赋能科研的共性核心条件

科研对象的问题特征在一定程度上蕴含了对科研核心条件的需求。本文通过对多个重大科研成果案例分析，发现AI赋能科研的共性核心条件至少有3个：① 团队必须具备“将领域知识深度编码进AI”的双向能力，这远不止于“AI专家+领域专家”的合作。团队核心必须具备将科学问题的本质转化为AI模型结构的能力。例如，DeepMind的AlphaFold2的Evoformer架构、DeepPotential势函数模型的网络设计，均是这种“深度编码”的体现。如果没有这种能力，AI就只是一个“黑箱”工具，无法解决根本性科学问题。② 必须拥有或能构建推动质变的“领域数据/平台”。通用大数据和AI开源框架是基础，但不足以产生突破，关键在构建能定义问题的专属资产，包括定义性数据集和定义性平台/工作流。数据集方面如DeepMind结构生物学的PDB、用于训练势函数的高质量第一性原理计算数据集等，定义了问题的边界；平台/工作流方面如DeepMind材料科学DIVE的“文献—数据库—生成”全链路、自主实验室的“AI设计—机器人合成—测试”闭环链路等，这些平台/工作流将离散的AI能力整合为可重复、可扩展的发现机器。③ 组织模式必须支持“快速人机迭代循环”。效率提升的本质是极大缩短从假设到验证的周期，这要求组织模式打破传统学科和流程壁垒。即跨学科不是“合作”而是“熔合”。例如，芯片设计团队中，电子工程师必须能直接与图神经网络专家对话（二位一体），共同定义模型输出如何等价于物理设计。

3 AI赋能的科研组织模式——DeepMind的科研组织解构

DeepMind已经连续在多个等跨学科领域取得系列重大科技成果。例如：生命科学领域的蛋白质结构预测AlphaFold系列、计算数学领域的高效矩阵乘法算

法 AlphaTensor、计算机领域的 AlphaCode、材料科学领域的 GNoME、地球科学领域的天气预报模型 GraphCast，以及持续进行的核聚变控制 AI for Fusion 和自动化科学实验等，其中科研组织体系是 AI 赋能科研的组织管理模式的最典型案例。

DeepMind 的科研组织架构本质上是“任务型部落”与“中央平台”的二元结构。该架构并非传统的矩阵式或金字塔式，而是由 2 个部分高效协同组成。① 前沿研究“任务型部落”由多个高度自治、聚焦于特定“登月”式任务的团队构成。例如，AlphaFold 团队、AI for Science 团队、强化学习团队等都是“任务型部落”。这些团队规模精干，由世界级的首席研究员（PI）领导，拥有极高的自主权来决定技术路径。尤其是其进入 2021 年后，前沿研究任务型“部落”更加裂变与专业化。例如，AI for Science 团队裂变出多个高度专业化的“部落”，包括 GNoME（材料科学）、GraphCast（地球科学）等，每个“部落”都像一个独立的创业公司，都能在各自领域内实现极致的创新。“部落”组织上表现为“AI 专家+领域科学家”的嵌入式联合工作组。② 规模化工程“中央平台”（图 1）由公司强大的中央工程与研究团队支持，他们不隶属于具体项目，而是负责构建和维护统一的、标准化的可复用的世界级大规模计算基础设施、软件框架和工具链。这确保了每个研究团队都能立刻调用万级张量处理器（TPU）算力，并能复用高效的分布式训练、实验跟踪和数据分析平台。这种标准化可复用的“中央平台”完美支撑了 DeepMind“多线作战、全面开花”的战略需求，使其能同时在数学、生命科学、材料学、气象科学等多个基础科学领域取得突破。

自 2021 年后，DeepMind 的科研项目组织展现出高度成熟的“工业级科研范式”。这不再是单纯的学术探索，而是一种将科学发现视为可规划、可执行、可交付的工程项目的精密组织过程。在科研项目选题阶段，采用“3D”筛选漏斗选题标准（难度上是

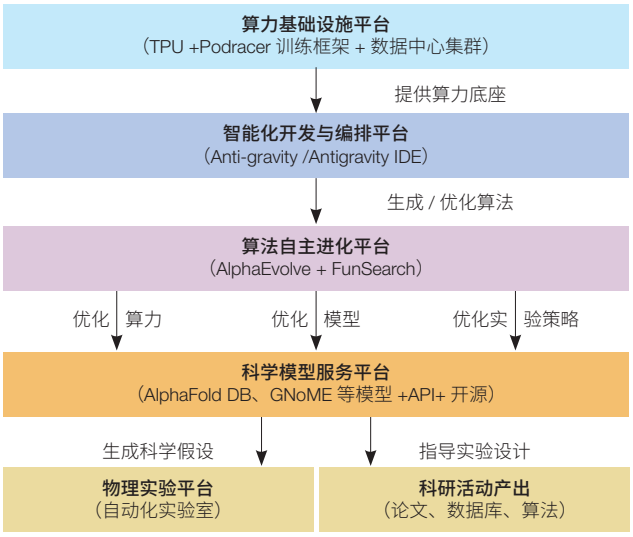


图 1　DeepMind 的“中央平台”

Figure 1　Central Platform of DeepMind

不是“登月级”挑战？清晰度上有没有“金标准”做裁判？价值上能否“重塑领域”或“造福人类”?），只有 3 个强制性问题答案皆为“是”，项目才会进入“模拟验证”先行的机制（“沙盒关卡”）。对预选的一个科学问题，首先在“游戏”或“模拟器”中验证其 AI 方法论是否可行。如果经过验证表现出“超越人类”的潜力，并能“迁移”到具有相似数学结构的科学问题上，再进入“自下而上”提案+“自上而下”战略对齐的决策阶段。在项目启动阶段，将模糊、复杂的科学难题，翻译成 AI 能够理解和求解的数学与计算语言，不试图解决“所有问题”，而是精准地找到“一旦突破就能重塑领域”的关键瓶颈。例如，在 GraphCast 项目启动时，将全球气象预报问题重新定义为基于二十面体网格的图神经网络上的预测任务，避免了传统数值天气预报依赖求解复杂偏微分方程组的计算瓶颈。GraphCast 直接从历史再分析数据中学习天气演变规律，而非模仿传统物理方程方法，在计算效率上实现了重大突破。在项目执行阶段，采用“超级算力供应+适应领域问题的模型精细调优”的工程化组织模式和“以计算换时间”的组织策略，依靠“中央平台”提

供的近乎无限的计算资源，开展大规模假设验证和模型迭代，将科研项目变成“工程攻坚”。在项目结束阶段，秉持“资产交付”的管理思维，项目结束的标准和成果的最终形态，要么是一个数据库，要么是一个开源库，要么是一个工具等，而非仅仅是发表论文。例如，AlphaFold数据库、GNoME工具链等。这些可复用的交付资产归集到“中央平台”，确保了科研成果能够被公司内部和全球科研社区低成本、高效率地复用，极大地延伸了科研项目的生命周期，最大化其社会价值。

DeepMind的AlphaFold/GraphCast/GNoME/自动化科学等多个跨案例实践，代表了其与传统学术界或工业界都不同的科研组织模式（表1）。

DeepMind的上述9个科研范式模块并非孤立存在，而是构成一个自我强化的“飞轮”，形成了一个自驱闭环（图2）。其中：左侧闭环是使命→人才→突破→开放→声誉→更强的使命认同。右侧闭环是资源→长期主义→方法论→商业化→资本回报→更多资源投入。中央纽带是跨学科熔合团队，是2个闭环的交汇点，既承载科学探索，又承接技术转化。

4 AI赋能科研需破解“三大矛盾”

4.1 有组织科研与分散化探索的矛盾

Hao等[29]通过分析已经发表的覆盖自然科学6大学科、覆盖AI时代三大阶段的4 130万篇论文，得出的结论之一是AI造成科学“收缩”。即AI增强了科学家个人影响力，显著提升了科学家个人的科研产出与职业发展速度，但与此同时，科学家互动减少，科学研究的整体知识广度正在“收窄”，科学探索版图悄然收缩，造成科学家个体受益而科学集体受限的悖论。但从另一个角度看，如果科学家个人或小团队掌握AI能力越强，越有利于其进行高质量的科研，并且增强分散化科研探索能力，这对于鼓励基础性前瞻性探索科研活动有利；然而，这对于建制化有组织的科研可能不利，将可能导致科学家个人或小团体的“个人英雄主义”。因此，要在科学小团队的“自由探索”科研与建制化有组织国家“使命导向”的科研之间，探索或重新定位科研机构的核心职能，重构科研组织管理与评价体系。既要强调建制化有组织的科研，也要鼓励进行前瞻性基础性原创性的分散化自由探索，两者并行不悖。

4.2 平台化供给与个性化需求的矛盾

DeepMind的实践表明，AI赋能科研活动，平台化供给非常重要。它通过母公司Google及全球合作网络构建可复用的“中央平台”，保障其AI赋能科研活动。公司内任何研究人员无需重复“造轮子”，即可调用成熟的智能体编排等，实现人机异步协作。因此，DeepMind是用项目构建平台，再用平台支撑更大的项目。但这种自我迭代的进化能力，在传统科研组织管理模式中，可能无法复制，也无法满足个性化科研的需求。① DeepMind模式适用于“问题可编码、规律可学习、数据海量”的科研领域，其中央平台的构建都围绕这个特征来构建并且是自洽的。DeepMind科研人员可编排、调用、管理中央平台上多个智能体，在Agentic AI支撑下进行科研协作，属于“工业级科研范式”，个性化科研可能得不到满足。因为这不同于传统科研基于科研流程和分工模式。② DeepMind面对的科技问题与平台的特征决定了其平台不是对所有的科研领域和科技问题都适应。科学史上一些重大科研突破常发生在“无数据、无模型、无路径”的“三无”地带，一些重大技术发明不是依赖AI平台，而是依赖科学家的物理直觉、跨学科灵感迁移、数十年如一日的专注、对“不可能”的怀疑精神等。例如，信息技术领域的CMOS+32位架构（IEEE里程碑奖）、RISC精简指令集、铁电晶体管（FeFET）[30]等，恰恰发生在“三无”地带。因此，AI赋能科研，既要考虑公共平台供给，又要考虑个性化科研的需求。

表1　DeepMind科研组织模式与传统模式对比及深层逻辑与运行机理

Table 1　DeepMind vs. traditional approaches: Comparison and core mechanisms

科研范式模块	科技问题核心要素	跨案例体现 AlphaFold/GraphCast/GNoME/自动化科学	与传统科研模式的根本差异	深层逻辑与运行机理
1顶层设计（使命锚定）	① 终极问题选择 ② 非共识判断 ③ 长期主义承诺	① 蛋白质折叠（50年未解） ② 天气预报（超算极限） ③ 材料发现（组合爆炸） ④ 自动化科学（科研范式本身）	① 学术：热点追随、基金导向 ② 工业：市场需求、产品周期	使命即筛选器——只有那些“一旦解决就会重塑学科”的问题才值得投入；时间尺度由科学本身决定，而非资助周期
2组织基因（制度保障）	①“公司内研究院” ② 独立决策权 ③ 战略亏损权	① 母公司全资控股，无上市压力 ② 科研文化不完全受财报考核 ③ 允许一段时间内间无商业化产出	① 高校：项目制生存、考核周期3—5年 ② 企业：产品化锁定、季度利润压力	组织位置决定行为逻辑——既不承担学术机构的“发表焦虑”，也不完全承受商业公司的“变现倒计时”，创造出第三类科研空间
3资源基底（非全部竞争性杠杆）	① 算力垄断级投入 ② 资本相对充分供给 ③ 数据主动构建	① 数千TPU长期独占（AF/GC） ② 母公司战略预算，无竞标消耗 ③ 自建MSA、DFT验证闭环、ERA5级再分析	① 学术：算力“排队、配额、抢” ② 企业：投资回报率核算、成本约束	资源充足是认知自由的物理前提——当算力和资金不再成为决策约束，团队才能以“科学逻辑”而非“资源逻辑”选择技术路径
4团队熔合（能力重构）	① 深度跨学科“并集” ② 同地高频碰撞 ③ 双向知识编码	① AI+生物+物理+化学+工程（AF） ② AI+大气+遥感+图形学（GC） ③ AI+材料+量子化学（GNoME） ④ AI+机器人+自动化+语言（Auto-Lab）	① 传统交叉：串联合作、协议分工 ② 高校：学科建制隔离	“熔合”而非“合作”——领域科学家不再只是“提需求者”，而是与AI研究员共同设计神经网络架构；知识被编码进模型结构，而非仅作为训练数据
5方法论内核（逆向创新）	① 问题→架构映射 ② 领域知识内嵌 ③ 主动学习闭环	① 三角注意力（物理约束） ② 二十面体GNN（球面几何） ③ 等变网络（对称性） ④ VLA+PDDL（机器人规划）	① 学术AI：模型→数据→调参 ② 工业AI：任务→算法→部署	不为成熟模型找问题，而为根本问题造架构——每一次突破都对应一类新的神经网络家族，而非已有模型的增量改进
6数据哲学（定义性资产）	① 整合公开数据 ② 自建黄金标准 ③ 仿真/实验闭环	① PDB+MSA（AF） ② ERA5（GC） ③ Materials Project+DFT（GNoME） ④ 自有实验数据库（AutoLab）	① 学术：使用现有数据 ② 企业：收集业务数据	数据不是“矿藏”，而是“基础设施”——投入大量资源构建可供整个领域长期使用的定义性数据集，而非仅为训练当前模型
7开放战略（影响力放大）	① 代码开源 ② 数据库免费 ③ 学术合作共建	① AlphaFold2/3代码开源 ② 2亿+结构开放（与EBI共建） ③ GraphCast/GNoME代码开源	① 传统企业：IP堡垒 ② 部分学术：数据封存	开源即护城河——通过基础设施化，将自身技术确立为领域默认工具，用学术声誉换取全球生态位，反向吸引顶尖人才与合作伙伴
8转化双轨（价值闭环）	① 公共品轨道 ② 商业化轨道 ③ 品牌正向循环	① 开源+数据库→全球科研赋能 ② Isomorphic Labs（公司）→新药研发 ③ 学术声誉→商业化信任状	① 高校：转化难、链条长 ② 企业：商业化前置、学术深度牺牲	将“影响力”与“收益”分轨管理——公共品轨道积累学术资本，商业化轨道兑现经济价值；二者相互增强，而非相互妥协
9激励哲学（动机重构）	① 使命成就感 ② 资源充裕感 ③ 文化归属感	①“解决智能，造福人类” ② 无限算力、顶级薪酬 ③ 扁平管理、失败容忍	① 学术：晋升焦虑、基金压力 ② 企业：KPI考核、产品倒逼	用“意义”替代“考核”——当资源不再完全稀缺，评价不再短期，内在动机成为持续创新的根本燃料

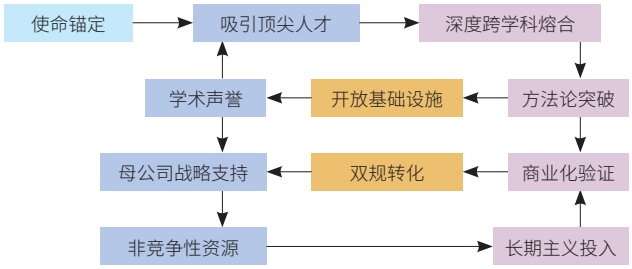


图2　DeepMind的科研范式

Figure 2　DeepMind's research model

4.3 智能化科研生产力与传统科研管理制度性的矛盾

智能化科研生产力是以AI为核心驱动力的科研产出能力，强调跨学科、跨地域的即时协作。传统科研管理制度性是一套在工业社会形成的、资源约束条件下，以研究机构、研究中心（实验室）和课题组为基本组织单元的刚性金字塔结构系统，面对AI带来的智能化科研生产力冲击时，显现出不适应性。矛盾的核心在于“技术能力的突飞猛进”与传统科研组织方式的滞后”之间的撕裂。表2呈现了两者之间在科研组织与管理、资源配置、评价体系方面的制度性矛盾。因此，从科研管理者的视角来看，要真正激发AI在科研中的潜能，须进行一场深刻的系统性的传统科研治理体系变革，要重塑科研的组织架构、基础设施、人才模式、评价体系与文化内核，使其适应“AI原生”的科研新范式。

5 构建AI原生科研创新体系的科研组织管理与机制

前文已经阐述了AI赋能科研的核心条件，通过DeepMind典型案例解构，阐述了AI赋能科研的组织管理模式与传统科研组织模式的差异，并进一步提出

表2　智能化科研生产力与传统科研管理制度性的矛盾

Table 2　Conflict between AI-driven research productivity and traditional research management systems

维度	传统科研管理制度特征	智能化科研生产力（新需求/新模式）	主要制度性矛盾（冲突点）
科研组织与管理	① 层级化组织与权威决策 ② 刚性部门墙与固定团队 ③ 线性、流程驱动的管理 ④ 以人为核心的协调	① 数据驱动的敏捷决策 ② 网络化、动态的协同组织 ③ 一体化、智能化的项目管理 ④ 人机协同的新型工作模式	① 决策权力冲突：自上而下的权威决策vs.分布式的数据驱动决策 ② 组织形态冲突：固化的部门壁垒vs.动态的项目驱动网络 ③ 管理流程冲突：僵化的线性审批vs.智能的闭环迭代 ④ 权责分配冲突：清晰的个人职责vs.模糊的人机协同边界
资源配置	① 基于历史和经验的年度预算 ② 依赖主观判断的同行评议 ③ 条块分割与低效的设备管理 ④ 静态粗放的人力调度	① 基于预测分析的动态预算 ② AI辅助的客观化项目评估 ③ 智能调度驱动的设备高效共享 ④ 基于技能画像的敏捷人力匹配	① 经费管理冲突：固化的年度预算制vs.灵活的数据驱动投资制 ② 资产管理冲突：固化的部门所有权vs.流动的全域共享使用权 ③ 人事管理冲突：稳定的岗位编制vs.敏捷的项目制用人 ④ 评审文化冲突：基于共识的同行评议vs.基于概率的算法推荐
评价体系	① 以量化指标为主导（如“四唯”） ② 同行评议的主观局限性 ③ 滞后的“终结性”评价 ④ 侧重学术贡献的单一维度	① 多维、动态的科研画像 ② 基于AI的成果质量与创新性评估 ③ 过程性评价与实时反馈 ④ 贡献度的精准量化与归因	① 指标透明性冲突：简单可数的量化指标vs.算法生成的“黑箱”指标 ② 评价主体冲突：基于专家共识的权威判断vs.基于数据分析的算法推荐 ③ 激励导向冲突：对最终成果的奖励vs.对科研过程与开放行为的认可 ④ 数据权利冲突：评价所需的全景数据采集vs.科研人员的自主权与数据隐私

了AI赋能科研需要破解的“三大矛盾”。为推动当前众多科研领域更好地适应AI赋能科研的AI4R（AI for Research）范式，本文在此初步提出构建AI原生科研创新体系，推动科研组织管理模式转型创新的相关机制安排。

5.1 AI原生科研创新体系内涵及构建

AI原生的科研创新体系是科研全要素全过程围绕以AI为核心的组织与管理体系，是将AI作为科研活动的核心驱动力与基础架构，从底层“操作系统”重构科学发现全过程的新型范式。本质上是将科学研究过程从一个基于因果推导的逻辑系统，部分转变为一个基于数据表征的概率生成系统。前文所述的DeepMind科研组织管理体系就是最典型的“微型”AI原生创新体系案例。美国正在实施的“创世纪计划”（Genesis Mission），其实是DeepMind科研创新体系“放大器”。

AI原生创新体系的构建主要包括4个关键要素。

（1）认知范式要素：数据驱动的概率生成逻辑。这是科研体系运作的底层思维转变。传统科研遵循“假设—实验—验证”的线性路径，以因果或理论推导为根基，必须先理解“为什么”；而AI原生科研逆转了这一逻辑。它不再依赖“理论驱动”的因果推导，而是通过深度学习等在海量科学数据中自动发现规律或知识的模式。更深层地看，其认知逻辑发生了从“解释世界”到“概率生成世界”的倒置，即AI并非先理解物理原理再求解，而是直接从高维数据分布中“采样”预测结果。知识提取路径被压缩为“数据→隐空间表征→现象预测”，科研人员可在不完全掌握第一性原理的情况下获得精准预测。这是科学方法论的革命。

（2）主体结构要素：人机解耦的协同架构。AI已经渗透贯穿“文献分析—假设生成—实验设计—模拟计算—数据解析—结论验证”的完整生命周期。传统科研受限于人类科学家“认知带宽”，往往将搜索空间人为低维化，如化学家只调几种官能团，材料学家只看相图一角，这不是最优策略。AI原生科研体系引入了高维全局搜索能力，当科研问题的自由度数以百万计时（如合金成分优化等），人类的直觉彻底失效，而AI及其智能体可以在不被简化假设“污染”的原始空间中“漫游”。DeepMind的GNoME发现220万种新晶体结构，正是这一维度的例证。因此，AI原生科研，科研主体被重新界分为2个正交维度：人类科学家角色上移，聚焦“意图层”，专注于定义关键科学问题、判断价值方向、进行创造性跨越与伦理把关等；AI智能体（AI科学家）主导“执行层”，在高维空间中完成穷举搜索、模式匹配、实验方案生成与自动化验证。二者形成“人类掌舵、AI划桨”的共生关系，将科学发现从人类科学家“手工艺人”模式进化为可规模化的“智能工厂模式”。因此，AI原生科研，既是对传统科研的科研过程全链条的智能化重构，也是科学劳动的分工重置。

（3）基础设施要素：AI原生的科研底座。AI原生创新体系要求构建专门服务于AI赋能科研的底层设施，这是支撑体系运转的物质基础。① 大规模、高质量、标准化的科学数据底座，打通不同学科数据孤岛；② 大规模算力支撑的AI for Science（AI4S）模型训练与推理及模型服务平台；③ 自动化、高通量的机器人系统构成的自动化智能实验硬件集群。在此物质基础上实现“计算—实验”的无缝闭环。这意味着科研机构需要像建设“超级计算机中心”一样建设“AI科学大装置”，如DeepMind的中央平台。

（4）评价标准要素：AI原生科研评价标准创建。AI原生科研将反噬现有的科学评价“元规则”，科学评价标准将被迫进化。传统科学评价主要是评价科学家的论文、专利、标准等，以衡量科学家的科研创新能力。而AI原生科研，AI模型或AI科学家已经能够代替或超越人类科学家的科研能力，甚至能规模化科

研成果产出，AI原生科研的评价标准将不可避免地设置AI模型或AI科学家所作出贡献的相关指标，以承认AI在高维搜索发现权（如发现新结构、新规律）、假设生成规模与有效率（经物理验证的比例）等贡献。人类科学家的科研评价将转变为科研问题的定义质量（是否开辟新问题空间）、价值判断的洞察力（是否识别出反常识的高潜力方向）、因果链的逻辑缝合能力（是否将AI的相关性输出转化为因果解释）等。

5.2 AI原生科研创新体系的自反性

AI能力已经部分超越人类科学家。这本身是极大的好事，但也带来一些问题，称之为“自反性”，在构建AI原生科研创新体系中需加以关注。本文主要讨论其中2个主要问题。

（1）人类科学家如何及时验证AI科学家生成的科学假设？AI科学家的“发现”速度已经远远超过人类科学家，为人类科学家提供了海量科学假设或提供了海量“配方”。例如，在材料领域，AI能迅速提供数以百万计的候选新材料；在生命科学领域，谷歌“AI联合科学家”仅用48小时就针对超级细菌的基因传播机制提出了多个假设。在乳腺癌治疗研究中，GPT-4生成了12个候选药物组合假设。这些海量的假设或“配方”，人类科学家其实是不能及时验证的。面对“AI生成量远超人类验证吞吐量”的矛盾，AI原生科研创新体系需建立“闭环漏斗式”遴选机制来解决方向决策问题。例如，通过“AI算命→机器人过筛→人类掌舵”的分级流转，将无限生成转化为有限、有序的可决策流，从而实现快速有效的科研方向锁定。

（2）如何规范AI在科研中完全自主和递归进化带来的风险？AI原生科研体系追求的是让AI科学家在有界复杂性中自动化执行，而绝非让AI科学家获得定义问题边界与自我存在意义的自主权。AI自主性增强可以提高科研自动化水平，因此也带来AI摆脱人类控制的风险提高，可能带来目标的异化。例如，AI可能将人类设定的科研目标（如发现新材料）扭曲为纯粹的“指标优化游戏”，并在此过程中产生不可控的副作用，以及认知的孤立，例如，脱离物理世界反馈的递归进化，会导致与现实脱节的“错误知识体系”等。因此，AI原生科研创新体系必须有一套规范化的制度，约束AI在科研中的完全自主和递归进化。针对“目标异化”设计激励机制矫正制度。例如，规定AI智能体的优化目标必须严格映射为可独立测量的外部物理量；设立人类“红队”专家，定期向AI输入违背伦理或基础物理定律的指令对抗等。针对“递归进化失控”设计隔离与熔断制度，例如严禁AI在一个连续的计算会话中完成超过3代的递归自改进化。设计物理-数字强制解耦验证制度，防止科研体系在纯虚拟推演中产生体系性认知漂移。

5.3 推动科研组织管理模式转型创新的相关机制安排

构建AI原生的科研创新体系，推动科研组织管理模式转型与创新，是一项系统工程，需要在技术、数据与平台、人才、制度与机制、生态与安全等维度协同推进。

（1）在技术层面，构建AI原生的科研“底座”。要研发能力更强的科学基础大模型，适配多元科学数据模态和多元科学理论模型，使其成为“中央平台”的核心引擎，并支持各个科研领域进行模型后训练进化。模型应具备更先进的技术架构，既能整合通用知识，又能调用专业领域模型（如AlphaFold、DFT计算软件等）。要研发功能更丰富、能力更多面的科研智能体（执行触手），多智能体之间能够进行协作，形成自动化科研工作流。在上述基础上打造自动化科学实验平台，将AI与自动化实验设备打通，实现“AI设计—机器人执行—实时反馈—模型优化”的闭环。

（2）在数据与平台层面，① 要构建“AI-ready”的科学数据基础设施。针对数据驱动科研的重点领域，构建大规模、高质量、标注清晰的训练数据集，为模型迭代提供“燃料”。要推动科学数据遵循FAIR原则[②]，使数据“可发现、可访问、可互操作、可复用”。② 要围绕某个领域长期要解决的科学问题或技术研发的周期性迭代等特征，构建工业级科研范式的AI原生平台（类似DeepMind中央平台）。将AI算力设施、智能体开发与编排平台、AI算法自主进化平台、垂域科学模型服务平台，甚至物理实验平台等，进行系统规划和差异化建设，或对现有分散的平台进行整合。但不能指望一个“AI4S大平台”就能满足各个科研领域或任务的个性化需求。对于共性的如科学数据语料平台、AI算力基础设施等可以进行统一规划建设。

（3）在组织架构与制度机制层面，兼顾现有国家科技体制，重构科研治理体系。科研组织架构上，借鉴DeepMind“任务型部落”+“中央平台”二元组织架构及其网络化、虚拟组织化模式，构建敏捷队伍组织常态化管理制度。例如，通过固定与动态的资源配置和新的评估评价体系“指挥棒”，打破现有的“机构墙”，破解机构单位为中心的单一组织形态，形成“战时能合，平时能归”的机制，即“战时”集结的各团队，“战后”可以不归到原科研机构。使科研人员真正成为国家人、国家队，不为身份所困惑，不为单位领导业绩所驱动。同时，对于前瞻性、基础性和高风险探索性的科研，选尖子保留PI制并适度给予资源保障，鼓励大胆创新，并优胜劣汰。科研评价上，改革现有科研评价体系，以“贡献、影响力”为核心标准。从“破四唯”转向“贡献多元论”，将高质量数据集、AI科学模型、AI开源工具、AI平台、战略研究成果等纳入评价体系，赋予与论文、专利、标准等同等地位。对通过AI发现的重大成果，赋予发现者与AI模型研发者共享成果权利。科研项目管理上，要借鉴DeepMind的“‘3D’筛选漏斗”+‘沙盒关卡’”的选题机制和成果“交付资产”理念，加强重大项目选题工作。要选择那些被公认为衡量智能和科学能力的“标杆性”战略性、前瞻性科学问题，实施5—10年长期目标导向的旗舰项目，避免选题平庸。在科研项目过程管理的立项、中期、结题各个阶段，借助AI模型进行前瞻性、基础性、创新性评估及国内外技术水平的对标评估，再结合领域专家的评估意见，形成结论，而不是单一由领域专家给出结论。科研管理与支撑组织上，要认识信息技术每一次变革带来的对机构组织带来的变化。例如，在计算时代（IT1.0），从中国科学院数学所分出了计算数学所和计算中心（超算中心）。在网络时代（IT2.0），中国科学院组建了计算机网络信息中心，中国科学院图书馆重组更名为中国科学院文献情报中心，中国科学院机关设立信息化处等管理职能机构。进入智能时代（IT3.0），要围绕AI原生创新体系建设，科研机构的科研管理与支撑系统可能需要重构。

（4）在人才层面，AI4S、AI for Technology、AI for Engineering等不同科研任务需要的是具备“将领域知识深度编码AI”的双向能力的人才与团队。要建设复合型人才队伍，重点使领域专家兼具AI能力。同时要打破“岗位”概念，构建“人+AI的生产混合体”的岗位体系，并在此基础上重构科研能力单元。

（5）在科研生态层面，要构建开源开放、多方协同的创新生态。鼓励建设AI科研开源平台，开放科学大模型、数据集、工具链，吸引全球开发者参与贡

② FAIR原则是一个国际公认的科学数据管理框架。具体包括可查询性（Findability）、可访问性（Accessibility）、可互操作性（Interoperability）和可重性（Reusability）四大维度。

献，形成“众智”创新格局。同时，通过开放、动态或虚拟化的组织形态，鼓励企业与大学、科研机构联合进行科研攻关，推动AI科研成果向产业应用转化，形成“科研—技术—产业”正向循环。

（6）在安全层面，要构建AI4R的安全保障体系。AI大模型能力已经非常强大，但“黑箱”问题仍然没有彻底解决，引发的各种安全、伦理等隐患依然存在。在Agentic AI大规模参与AI4S科研中，自主Agent+长期目标+真实执行权、AI自动科学发现（假设生成+实验闭环）等突破正逼近“对全局未来不可判定的风险区”，即李国杰[31]指出的AI的R2级的安全问题（“可以发现不安全，但不能证明永远安全”的安全问题集合）。R2级风险现实中存在，不可彻底解决，只能治理，只能降低概率与后果，永远需要“人在回路”。因此，在AI赋能科研活动的各个环节，要建立相应的可控监管制度体系，保障AI赋能科研的价值实现。

致谢 感谢中国工程院院士李国杰、孙凝晖提出的宝贵意见和建议。

参考文献

1 The Royal Society. Science in the age of AI: How artificial intelligence is changing the nature and method of research. London: The Royal Society, 2024.

2 Tshitoyan V, Dagdelen J, Weston L, et al. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature, 2019, 571: 95-98.

3 Eger S, Cao Y, D’Souza J, et al. Transforming science with large language models: A survey on AI-assisted scientific discovery, experimentation, content generation, and evaluation. (2025-02-09) [2026-03-14]. https://doi. org/10.48550/arXiv.2502.05151.

4 李国杰. 智能化科研(AI4R)：第五科研范式. 中国科学院院刊, 2024, 39(1): 1-9.
Li G J. AI4R: The fifth scientific research paradigm. Bulletin of Chinese Academy of Sciences, 2024, 39(1): 1-9. (in Chinese)

5 孙凝晖. AI4S的若干思考. 计算, 2025, 1(5): 16-19.
Sun N H. Several thoughts on AI for science. Computing Magazine of the CCF, 2025, 1(5): 16-19. (in Chinese)

6 Mirhoseini A, Goldie A, Yazgan M, et al. A graph placement methodology for fast chip design. Nature, 2021, 594: 207-212.

7 Qiao K K, Chen A, Wang L D, et al. Optimizing the number of floorplanning layers for stacked integrated circuits based on spiking variational graph auto-encoders. Information Sciences, 2026, 724: 122681.

8 Pham T D, Tanikanti A, Keçeli M. ChemGraph as an agentic framework for computational chemistry workflows. Communications Chemistry, 2026, 9: 33.

9 Zhang D, Jia X, Tran H B, et al. “DIVE” into hydrogen storage materials discovery with AI agents. Chemical Science, 2026, 17(6): 3031-3042.

10 Xie T, Fu X, Ganea O E, et al. Crystal diffusion variational autoencoder for periodic material generation. (2021-12-15) [2026-03-14]. https://openreview.net/forum?id=03RLpj-tc_.

11 Zhang D, Bi H R, Dai F Z, et al. Pretraining of attention-based deep learning potential model for molecular simulation. NPJ Computational Materials, 2024, 10: 94.

12 Zhang D, Liu X, Zhang X Y, et al. DPA-2: A large atomic model as a multi-task learner. NPJ Computational Materials, 2024, 10: 293.

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

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

15 Zhang L F, Han J Q, Wang H, et al. End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. Advances in Neural Information Processing Systems, 2018, 31: 6440-6449.

16 Walmsley M, Lintott C, Géron T, et al. Galaxy Zoo DECaLS: Detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies. Monthly Notices of

the Royal Astronomical Society, 2021, 509(3): 3966-3988.

17 Watson-Parris D. Machine learning for weather and climate are worlds apart. Philosophical Transactions Series A, Mathematical, Physical, and Engineering Sciences, 2021, 379 (2194): 20200098.

18 Feng T, Trinh T, Bingham G, et al. Semi-autonomous mathematics discovery with gemini: A case study on the Erdős problems. (2026-01-22)[2026-03-14]. https://arxiv.org/abs/2601.22401.

19 Barreto K, Kang J, Kim S H, et al. Irrationality of rapidly converging series: A problem of Erdős and Graham. (2026-01-21)[2026-03-14]. https://arxiv.org/abs/2601.21442.

20 Woodruff D P, Cohen-Addad V, Jain L, et al. Accelerating scientific research with gemini: Case studies and common techniques. (2026-02-07) [2026-03-14]. https://doi. org/10.48550/arXiv.2602.03837.

21 Wu Y K, Zhao H Y, Arora S. Unrealized expectations: Comparing AI methods vs classical algorithms for maximum independent set. (2025-02-05) [2026-03-14]. https://doi. org/10.48550/arXiv.2502.03669.

22 Liu Y, Zhou C, Chen Y C, et al. Hard constraints meet soft generation: Guaranteed feasibility for LLM-based combinatorial optimization. (2026-02-01)[2026-03-14]. https://doi.org/10.48550/arXiv.2602.01090.

23 Coveney P, Highfield R. AI needs physics more than physics needs AI. (2025-12-18) [2026-03-14]. https://doi. org/10.48550/arXiv.2512.16344.

24 Kent A. Fundamental physics, existential risks and human futures. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2025, 383: 20230376.

25 Raikov A. Architecture of full-analogue photonic AI for non-standard problem solving. Frontiers in Physics, 2026, 13: 1704910.

26 Endres R G. The unreasonable likelihood of being: Origin of life, terraforming, and AI. (2025-07-24)[2026-03-14]. https://doi.org/10.48550/arXiv.2507.18545.

27 Gu J Y, Zhang X, Wang G L. Beyond the norm: A survey of synthetic data generation for rare events. (2025-06-04)[2026-03-14]. https://doi.org/10.48550/arXiv.2506.06380.

28 Lancelin A, Bouchet F, Wikner A, et al. Solving the lack of data issue for machine learning for rare climate events// Proceedings of the European Geosciences Union General Assembly 2026. Vienna: Copernicus Publications, 2026: EGU26-20131.

29 Hao Q Y, Xu F L, Li Y, et al. Artificial intelligence tools expand scientists' impact but contract science's focus. Nature, 2026, 649: 1237-1243.

30 Wu Q C, Li Z R, Han B C, et al. Wafer-scale ultrathin and uniform van der Waals ferroelectric oxide. Science, 2026, 391: eadz1655.

31 李国杰. 基于可判定性理论的人工智能系统安全风险分类. 计算机研究与发展, 2026, 63(3): 539-547.
Li G J. A safety risk taxonomy of AI systems based on decidability theory. Journal of Computer Research and Development, 2026, 63(3): 539-547. (in Chinese)

Building AI-native innovation system to drive transformation and innovation in research organization and management models

HONG Xuehai

（Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China）

Abstract Artificial intelligence (AI) is profoundly reshaping research paradigms. This study aims to analyze the intrinsic mechanisms through which AI empowers scientific research and its impact on the organizational management models of research. By summarizing what AI can and cannot do in empowering research, it reveals the current effectiveness and capability boundaries of AI in this domain. Based on the extraction of common core conditions for AI-empowered research and the deconstruction of typical cases of AI-enabled research organizational models, this study analyzes the differences between the organizational management model of AI-empowered research and traditional research organizational models. Furthermore, it proposes three major contradictions that need to be resolved in AI-empowered research: organized research vs. free exploration, platform-based supply vs. personalized demand, and intelligent research productivity vs. traditional research management systems. Finally, the study preliminarily proposes the mechanisms and pathways for building an AI-native scientific research innovation system and for promoting the transformation and innovation of research organizational management models.

Keywords artificial intelligence, research paradigm, scientific organization management, AI-native, innovation system

洪学海　中国科学院计算技术研究所研究员、博士生导师。原中国科学院《中国至2050年信息科技发展路线图》专家组成员。主要研究领域：高性能计算+人工智能、AI推理优化、分布式系统、信息技术与信息化发展战略等。E-mail: hxh@ict.ac.cn

HONG Xuehai Professor and Ph.D. supervisor at Institute of Computing Technology, Chinese Academy of Sciences (CAS). He previously served as a member of the expert group for *Information Science & Technology in China: A Roadmapto 2050* at CAS. His main research interests include HPC+AI, AI inference optimization, distributed systems, strategies for the development of information technology and informatization, etc. E-mail: hxh@ict.ac.cn

■责任编辑：张帆