

本刊专稿

人工智能赋能科学知识发现的范式变革与生态重构

王从余¹, 米浩源¹, 戚文秀², 程鹏³, 张东晓^{4*}

摘要 人工智能赋能科学(AI for Science, AI4S)正通过整合科学先验知识与数据驱动方法,推动科学探索范式的变革。其中, AI4S 知识发现已成为国家战略层面的关键议题,旨在分析其概念边界、方法谱系及协作机制,并探讨其对科研生态可能产生的影响。首先,概述 AI4S 的知识嵌入与知识发现 2 条互补路径及其形成的闭环,并指出 AI4S 知识发现的核心贡献在于归纳自动化与知识显式化,但这一贡献更集中于现象与经验定律层面,向构造性理论的跃迁仍依赖人类的机制解释与验证。继而,基于表示—搜索—评估(representation—search—evaluation, RSE)三元框架,揭示 AI4S 知识发现方法的差异主要在于表示空间的开放程度,搜索与评估则围绕此开放度协同发展。在此基础上,明确人机协作中人类的核心职责是划定认识目标、提供语义约束和主导价值裁决。最后,讨论 AI4S 知识发现对科研生态的可能影响:在数据密度、自动化与验证条件较成熟的领域,科学家角色可能从任务执行者向“元认知中枢”转变;科研协作也可能呈现出组内更扁平、组际更分化的趋势,这种分化有望通过平台化资产、统一接口及有效的验证治理机制转化为更高质量的跨组协作起点。

关键词 人工智能; AI4S; 知识发现; 人机协作

美国“创世纪计划”(Genesis Mission)以比肩“曼哈顿工程”(Manhattan Project)的战略雄心开启人工智能变革科学研究范式的国家行动,目标是“在未来 10 年内使美国科学与工程的生产力和影响力翻倍”^[1]。这场行动的核心正是人工智能赋能科学(AI for science, AI4S)的知识发现。AI4S 知识发现是指在科学问题被形式化为可计算对象之后,利用机器学习与符号推理等方法,在数据与先验约束的交互中自动生

成可形式化表达的规律,并使其能够被解释、检验与迁移。其价值不仅在于提升科研生产力和影响力,更在于改变人类探索未知、总结规律的方式。围绕 AI4S 知识发现,系统梳理其理论与实践边界、核心贡献与范式演进;在此基础上,界定 AI4S 知识发现中人类科学家的核心职责,探讨其对科研生态的微观、中观及宏观层面影响,以期理解人工智能时代科学知识发现的发展规律、推动 AI4S 知识发现健康有序发展提供理论参考。创新点在于:提出 RSE 三元框架下的 AI4S 知识发现方法脉络及人机协作机制,并据此对科研协作与组织结构的可能变化作出讨论与展望,为 AI4S 知识发现的健康发展提供治理思路。

1. 中国人民公安大学侦查学院, 北京 100038

2. 北京大学哲学系, 北京 100871

3. 中国科普研究所, 北京 100081

4. 宁波东方理工大学, 宁波 315200

收稿日期: 2026-03-03; 修回日期: 2026-06-30

基金项目: 国家社会科学基金一般项目(25BKX031)

作者简介: 王从余, 副教授, 研究方向为人工智能的社会心理影响, 电子信箱: wangcongyu@ppsuc.edu.cn; 张东晓(通信作者), 教授, 美国国家工程院院士, 研究方向为科学机器学习, 电子信箱: dzhang@eitech.edu.cn

引用格式: 王从余, 米浩源, 戚文秀, 等. 人工智能赋能科学知识发现的范式变革与生态重构[J]. 科技导报, 2026, 44(14): 23-36; doi:10.3981/j.issn.1000-7857.2026.03.00009

1 AI4S 的双重路径:从知识嵌入到知识发现的互补闭环

近年来,人工智能与科学研究的深度融合,逐渐形成以 AI4S 为代表的新兴研究范式。从 AI 与 science 的交互关系来理解, AI4S 的本质并不是用 AI 替代科学,而是让科学先验和数据驱动形成双向耦合。基于此,当前研究大致沿着 2 条相对但互补的路径展开。第 1 条路径是知识嵌入(knowledge embedding),即将科学先验知识(如守恒律、对称性与不变性、结构约束与量纲约束等)直接嵌入机器学习模型之中。知识嵌入的典型代表是物理信息神经网络(physics-informed neural networks, PINNs),其是一种结合了深度学习和物理学知识的机器学习模型。通过把物理知识写进模型,既能借助机器学习的强拟合能力去刻画高维、复杂、非线性的映射关系,从而提升预测精度,又能显著增强模型的可解释性、泛化性与物理合理性^[2-3]。但知识嵌入本质上仍是在既有科学先验框架内,提升模型的预测性能、泛化能力与物理一致性。第 2 条路径是知识发现(knowledge discovery),即从观测数据或数值仿真数据中,自动提炼可形式化表达的规律(如控制方程、关系式或函数封闭项等),从而生成新的、可解释且可迁移的科学知

识^[4-5],此类研究通常不是依赖单一算法,而是将不同技术环节组合起来:神经网络可用于构建连续状态场或代理模型,以缓解稀疏、含噪数据下的导数估计问题;稀疏回归、遗传算法、强化学习或大语言模型生成等方法,则用于在候选表达空间中搜索可能的方程结构;最终再依据拟合误差、简洁性、物理一致性和数学易解性等标准筛选候选规律。知识嵌入强调利用(exploitation)已有科学知识提高数据处理效率和可信度;知识发现强调从原始数据中探索(exploration)新的知识, AI4S 借助这 2 条路径的科学验证闭环重塑科学研究的流程,用知识约束模型、用数据生成新知识,再用新知识升级约束,从而给出既能预测、又能解释、还能被检验的新知识或新机制。综上, AI4S 的范畴很广,其本质是科学先验与数据驱动的双向耦合,当前研究包含知识嵌入与知识发现 2 条互补路径,本文主要探讨 AI4S 知识发现,其核心概念及其相互关系如图 1 所示。

2 AI4S 知识发现的核心贡献:归纳自动化与知识显式化

在大科学时代,科学观测数据迎来快速增长,然而从数据中提炼规律、探明机制、形成理论的速度并

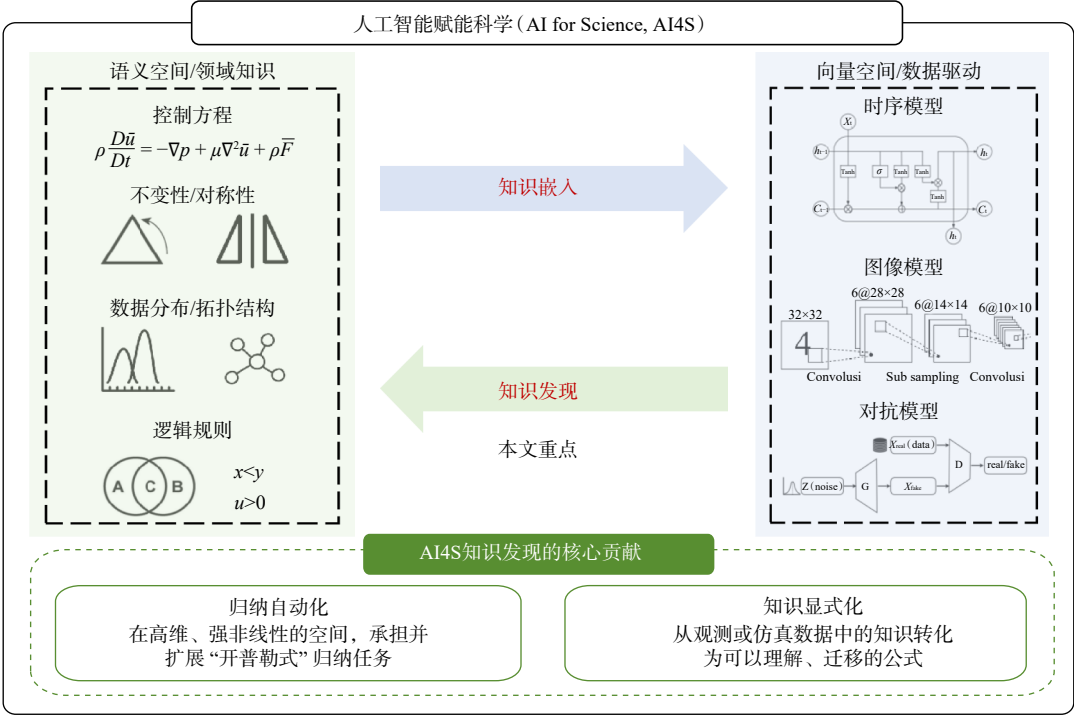


图 1 概念框架图

未能实现同步提升, AI4S 知识发现可以更为高效地将科学数据转变为可解释、可迁移的科学规律, 从而加速科学研究进展, 因此, 近年来受到更加广泛的关注^[6]。

2.1 归纳自动化

从更宏观的科学史视角看, 打通连续向量空间与离散语义空间之间的壁垒, 并非人工智能时代才出现的新问题, 而是根植于科学发现(本文将科学发现限定为科学知识的增量生成, 而不直接纳入仪器革新、实验方法改进等其他科学实践要素)的基本范式之中。科学发现长期以来依赖 2 种互补的认知模式, 即从观测或实验数据中总结规律的经验归纳(induction)与基于已有规律进行推演与统一解释的理论演绎(deduction)。一个典型的案例来自天文学。16 世纪, 第谷·布拉赫(Tycho Brahe)在没有望远镜的条件下, 持续观测并精确记录了长达 30 年的火星运动数据, 形成了当时极为罕见的高精度观测数据集。在此基础上, 他的学生开普勒又投入了近 18 年的时间, 几乎完全依靠人类理性与直觉, 对观测数据进行反复比对、修正与重组, 归纳出了行星运动三定律, 包括著名的椭圆轨道定律。这是典型的从数据到规律的经验归纳过程。随后, 牛顿在此基础上通过严格的数学推导提出万有引力定律, 实现了对行星运动现象的统一解释^[7]。这一历程表明, 从数据中发现规律的经验归纳在科学发现中具有基础性地位, 而定律能否进一步上升为理论, 则取决于后续的演绎整合与机制说明。

从知识生产范式来看, AI4S 知识发现可承担并扩展“开普勒式”归纳任务。人类大脑有认知瓶颈, 难以想象高维、强非线性的空间, 因此, 无法高效处理高维空间中的复杂关系; 而计算机擅长处理高维、强非线性数据结构, 能在稀疏、高噪数据条件下高效搜索潜在方程形式, 进而突破人类从数据中归纳规律的认识瓶颈^[8]。综上, AI4S 知识发现的贡献之一是归纳自动化, 是对传统科学发现范式的一种方法论回应。

2.2 知识显式化

AI4S 知识发现作为知识生产新范式, 并不必然每次都带来全新定律或颠覆性理论, 更常见也更具基础意义的贡献是知识显式化, 其目标是从稀疏、含噪的观测或仿真数据中, 将传统 AI 黑盒模型中只能表示(有一些不能表示)的知识转化为可以理解、迁移的公式, 如常微分方程(ordinary differential equation,

ODE)、偏微分方程(partial differential equation, PDE)及其闭合/本构关系、代数关系式, 以及参数化模型(如湍流亚格子应力参数化)等。其中, AI 驱动的 PDE 发现被视为 AI4S 知识发现中最具代表性的基础任务之一。PDE 构成了连接经验现象与机制解释的核心表达载体, 从欧拉方程、波动方程和热传导方程, 到纳维-斯托克斯(Navier-Stokes)方程及更复杂的现代方程体系, 科学理论的演进在很大程度上可以被理解为一个用 PDE 刻画自然现象的过程。

事实上, 科学知识体系的构建高度依赖公式、变量关系和可推演的数学结构。一个方程或结构模型的自由参数可以少到个位数。这种低参数、强结构的形式, 天然带来更高的计算效率, 也减少了对数据规模的依赖, 降低了过拟合的空间, 往往具有更好的泛化能力。而大规模神经网络具备极高的预测精度, 却往往以百万级, 甚至更高的参数量来吸收数据模式。如果 AI 能从数据中找出一个方程, 不仅是在把规律解释清楚, 而且是在用更经济的方式表达, 使之更可迁移、更顺畅地融入现有科学知识体系^[9]。这种人类可理解的表征还促使科学家能够进行进一步的逻辑推导, 甚至被启发新的灵感。

现有研究已在多个典型场景中展示了这一贡献。例如, 在风机尾流建模中, 符号回归把动态规律显式化为少量参数的解析表达, 使其更易跨工况迁移并接入既有尾流建模体系^[10]。在黏性重力流问题中, 基于高分辨率微观仿真数据的方程发现揭示了短期动力学中一个此前未被显式纳入宏观模型的补偿项, 从而修正传统模型在短时间尺度上对黏性效应的系统性低估^[11]。在地形—降水关系建模中, 通过 AI 自动分析数据发现获得的可解释方程不仅提升了预测性能, 其空间分布的系数结构还揭示了不同区域的主导物理机制, 并支持气候因素与地形效应的结构性解耦^[12]。在湍流大涡模拟中, 数据驱动的参数化发现能够恢复更为紧凑的亚格子应力表达形式, 从而提升模型在不同分辨率条件下的泛化能力^[13]。

2.3 贡献边界辨析

AI4S 知识发现的瓶颈可分为 2 层: 一层是连续—离散的表示鸿沟; 另一层是从经验规律到理论解释的层级跃迁。首先, 知识发现的技术难点可以概括为, 如何把连续向量空间中的函数逼近与搜索能力,

稳定地转化为离散语义空间中的形式化知识表达。现代机器学习模型(尤其是深度神经网络)的核心是向量化,本质上运行于连续的向量空间,擅长高维表示学习与复杂函数逼近,但其内部表示通常缺乏明确语义。与之相对,科学知识往往存在于离散的语义空间之中,以各类物理、化学概念为基础,组成控制方程、不变性结构、拓扑关系或量纲约束等形式的领域知识。如何打通连续向量空间与离散语义空间的壁垒,不仅是算法性能问题,还叠加现实数据条件与方程结构的双重约束。一方面,实验或高保真仿真往往代价高昂,导致数据天然稀疏;观测和数值计算不可避免地引入显著噪声。另一方面,真实系统的控制方程常包含高阶导数、强非线性、复合结构甚至分式形式;方程系数既可能是常系数,也可能是可表达的变系数,甚至是不可解析的随机场,使候选空间既大且形态多样。即便得到若干候选表达式,也不能仅凭拟合精度决定优劣;还需要兼顾简洁性、稳定性与物理一致性,才能够被置入可解释、可检验的知识框架之中。只有同时兼顾简洁性、稳定性与物理一致性,才能从 AI 系统中获得真正可解释、可验证、可迁移的知识增量,并据此与知识嵌入形成可循环的闭环。

其次, AI4S 知识发现的优势更多地体现在现象描述与规律预测这 2 个层面。如图 2 所示,科学知识可分为 4 个层级:一是现象(phenomenon),即从杂乱数据中识别稳定、可重复的模式,是经验定律的素材;二是经验定律(empirical laws),即通过数据分析直接获得的知识,不引入额外变量,完成从现象描述到规律预测;三是构造性理论(constructive theories),它不止于从经验现象中抽取规律,而是提出具有明确机制指向的结构化模型,回答“为何如此”,例如,牛顿第二定律对受力运动的建模;四是原理性理论(principle theories),即关于理论的理论,为下层理论设定约束,

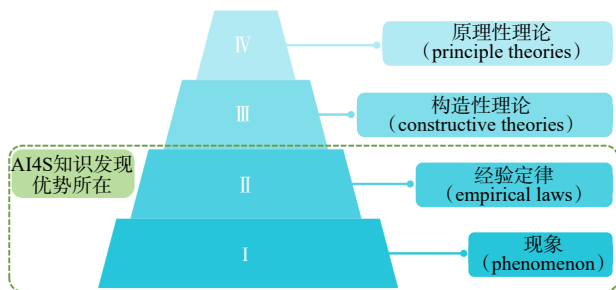


图 2 科学知识结构的 4 个层级

如相对性原理与绝对时空观之于牛顿力学^[14]。其中,经验定律与构造性理论的区分尤为关键,二者之别不在拟合是否精确,而在是否触及机制。据此, AI 产出的方程要从经验定律上升为构造性理论,大致需经历由浅入深的三重检验。一是机制可解释,即方程的各项与参数能对应到已知机制,而非仅为拟合而引入的自由参数。二是约束相容与可外推,即规律不仅在训练数据范围内成立,也与守恒律、对称性、量纲分析等既有约束相容,并在更广情形下依然有效。三是可推导,即该规律能被纳入更上层的原理,或反过来将已知定律作为特例推出。这三重恰恰不是 AI 当前的所长,仍主要依赖人类。

这一分工在自动化薄层色谱(thin-layer chromatography, TLC)研究中有清晰体现。该工作中,机器学习模型从数千条标准化数据中习得了“拓扑极性表面积越大、保留因子值(retardation factor, R_f)越小”这一稳定且可预测的结构—极性关系,在经验定律层面达到了高精度。但模型本身并不解释这一关系为何成立。研究者进一步从硅胶表面吸附平衡出发做出理论推导,证明 R_f 值是竞争吸附反应吉布斯自由能变化的 sigmoid 函数,为该经验规律提供了热力学机制层面的支撑^[15]。正是这一步由人类完成的机制解释,体现了经验定律向构造性理论跃迁时人类不可替代的作用。事实上,一些自动化合成平台虽能高效产出可重复的合成结果,其输出却主要停留在经验层面,“为何如此”仍有待人类补足。换言之, AI4S 知识发现更直接、更稳定的贡献,是将传统上高度依赖人类直觉的归纳过程计算化,并在稀疏、高噪的现实条件下,尽可能恢复可解释的形式化规律。而向更高理论层级的跃迁,则有赖于更完备的科学数据库、更强的跨尺度建模能力,以及人类在机制解释、实验验证与理论整合上的持续参与^[16-17]。

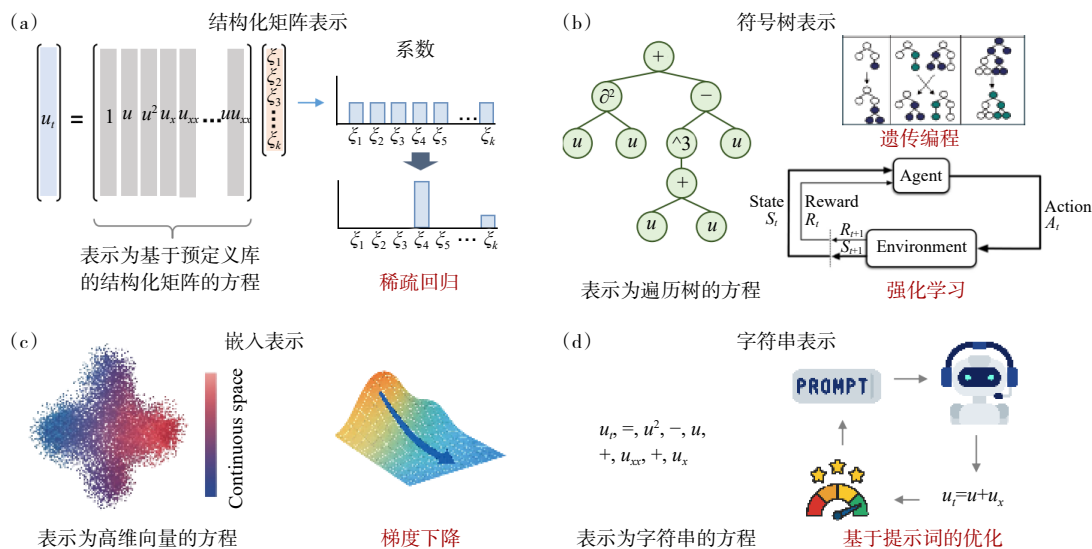
3 AI4S 知识发现的范式演进:以表示空间的逐步开放为主线

在明确变量选择、约束条件等科学问题的界定之后,现有 AI4S 知识发现方法可分解为 3 个相互关联但逻辑上可区分的核心环节:表示(representation)、搜索(search/optimization)与评估(evaluation),即 RSE 三

环节。RSE 并非对既有方法的简单重新命名,也不是对机器学习或符号回归一般流程的重复概括,而是面向 AI4S 知识发现任务的一种统一抽象。一般机器学习主要关注从输入数据到预测输出的优化过程,传统符号回归主要关注在表达式空间中搜索可解释函数;相比之下, AI4S 知识发现面对的是更具科学语义的知识对象。科学规律通常以方程、函数关系、闭合项或结构模型等形式存在,而计算机和 AI 模型难以直接处理自然书写的数理表达。因此,知识发现首先需要将这些科学表达转化为机器可计算、可搜索、可优化的表示对象,这正是表示环节(R)的核心作用。随着候选表示空间扩大,候选结构数量迅速增加,搜索难度随之上升,因此需要相应的搜索策略来生成、筛选和优化候选结构(S);同时,搜索过程还必须依赖评估环节获得反馈,因为 AI4S 知识发现不仅要找到“拟合数据”的表达式,更要判断什么是“好的科学方程”,因此,需要针对性的评估标准(E)。现有的判准包含 4 个层面:一是精确性,即方程对观测数据或数值解的拟合误差是否足够小;二是简洁性,强调在解释力相近时优先选择更简单的表达,以符合奥卡姆剃刀原则;三是物理一致性,即候选方程是否满足关键先验约束(如守恒、对称性、边界条件、量纲一致等)或在嵌入物理约束学习框架(如 PINNs)后能否带来一致性的误差改善;四是数学易解性,即方程在形式上是否更利于求解、分析与推导,如避免病态项、极端不

适应结构,或更便于稳定数值求解。简言之,一个挖掘出的方程模型应当平衡简约性和精准性,从而具备较好的可解释性和实用性^[18]。RSE 的增量在于将“科学表达—机器表示—候选搜索—科学验证”这一完整链条显式化。

在 RSE 框架中,表示是区分不同知识发现方法的起点。不同表示方式实际上定义了不同的候选方程空间。如图 3 所示,现有 AI4S 知识发现中的方程表示可按照表示空间的开放程度概括为 3 类:第 1 类是封闭候选集表示,通常将预设函数项写成结构化矩阵,并通过稀疏回归筛选关键项(图 3(a));第 2 类是可扩展候选集表示,即在预设变量、导数阶次、算子或基因片段的基础上,通过进化规则、组合规则或库扩展机制生成新的候选项;第 3 类是开放形式表示,包括符号树表示(图 3(b))、嵌入表示(图 3(c))和字符串表示(图 3(d)),分别将方程表达为由变量与算子组成的组合树、连续隐空间向量或可由语言模型处理的序列。本文对方法谱系的划分并不是以具体算法名称为标准,也不是将不同方法理解为线性替代的“三代”,而是以候选方程表示空间的开放程度为主线。由此, AI4S 知识发现方法的演进并不只是算法性能的提升,而首先体现为候选表示空间的逐步开放:从固定项库中的稀疏筛选,到候选项库的规则化扩展,再到开放形式方程结构的生成与搜索。需要强调的是,表示越开放,搜索越需要更强的生成与探



按表示空间的开放度, (a) 为封闭候选集表示(第 1 类), (b)、(c) 和 (d) 均为开放形式表示(第 3 类)

图 3 AI4S 知识发现研究中常见的 4 类方程表示

索能力, 评估也越需要在准确性之外兼顾其他目标, 以在更大的假设空间中守住科学可解释性与可检验性^[19]。因此, RSE 既可作为分析既有方法差异的坐标, 也可作为设计未来 AI4S 知识发现系统的框架: 先确定表示空间, 再匹配搜索策略, 最后构造能够反馈搜索过程并支持科学确认的评估机制。

如表 1 所示, 3 类方程表述方法均可纳入 RSE 框

架理解, 不同路线并非替代关系, 而是在不同先验条件、数据质量和任务复杂度下具有互补性。它们的核心差异首先发生在表示空间层面; 搜索策略和评估机制则围绕表示空间的开放程度进行配置。因而, RSE 不是以算法名称为中心的分类法, 而是以“机器如何表示规律、如何搜索规律、如何判断规律”为核心的问题分解框架。

表 1 3 代方程表示方法和 RSE 框架的对应关系

方法类型	R: 表示空间	S: 搜索策略	E: 评估验证	适用范围	优劣分析
封闭候选集	预设函数项库、结构化矩阵、稀疏系数	稀疏回归、逐步筛选、最小二乘	精确性、简洁性	适合先验较强、结构较明确的问题; 关键在于构造高质量候选库	可解释、稳定, 但难发现库外结构
可扩展候选集	预设变量、导数、算子或基因片段的组合扩展	遗传算法、进化搜索、库扩展	精确性、简洁性、物理一致性	适合候选结构不完全但仍有明确构造单元的问题; 关键在于扩展规则与搜索效率	比固定库更灵活, 但仍受预设片段和规则约束
开放形式表示	符号树、隐空间嵌入、字符串序列	遗传编程、强化学习、梯度优化、LLM 生成、工具调用	精确性、简洁性、物理一致性	适合结构高度未知的开放发现任务; 关键在于控制搜索空间并强化科学验证	表达能力强, 但搜索空间大, 需严格验证

3.1 第 1 代方法: 基于结构化矩阵表示的封闭候选集

第 1 代方法是结构化矩阵表示, 其核心思想是在预先给定的候选项库中, 通过稀疏约束筛选出少量关键项, 从而确定方程结构。具体做法通常是人工构造包含变量、多阶导数及其多项式或三角函数组合的字典项库, 然后通过稀疏回归(如 LASSO、岭回归等)求解非零系数项, 得到控制方程形式。这类表示的假设空间最为封闭, 但可解释性强, 计算代价相对可控, 在噪声较低、结构相对简单的问题中表现稳定^[20-22]。

然而, 其根本局限在于对先验候选集的强依赖。该类方法依赖预设候选项库, 因此, 其发现能力受到候选空间完备性的限制。但是当系统结构未知或高度复杂时, 合理候选项往往难以事先枚举, 从而引发先验依赖悖论: 当研究者能够写出足够完整的候选集时, 往往已经接近知道方程结构本身; 当真正探索未知问题时, 又难以给出完备的候选集。此外, 受制于稀疏回归算法的限制, 此类先验候选集的规模不能过大, 包含的候选项往往有几十项, 难以覆盖所有可能的场景。上述问题限制了该类方法在开放科学发现任务中的适用范围^[23-25]。

3.2 第 2 代方法: 基于库扩展与基因片段组合的可扩展候选集

第 2 代方法是可扩展候选集方法, 其核心是在预

设变量、导数阶次、算子或基因片段的基础上, 通过规则扩展、组合变换或进化操作生成新的候选项。与封闭候选集方法相比, 这类方法不再局限于静态候选库, 能够在已有构造单元之间产生新的交互项或复合项。例如, 真实方程中某个交互项即使不在初始候选集中, 只要其组成因子已被纳入编码空间, 算法仍可能通过组合与重排生成这一项^[5]。部分 PDE 发现方法会先使用神经网络拟合观测数据, 并通过自动微分获取高阶导数, 随后将变量、导数阶次及其乘法组合编码为“基因片段”, 再通过交叉、变异等遗传操作生成候选结构^[26]。由此看, 神经网络主要承担连续状态场拟合与导数估计功能, 遗传算法或进化搜索主要承担候选项扩展与结构搜索功能。方程选择一般同时考虑拟合误差与结构复杂度, 以平衡精度与简洁性。

然而, 现实中的动态非线性系统往往非常复杂, 其控制方程也经常包含分式结构、复合函数求导等多样化形式。对于这类复杂结构的函数项, 可扩展候选集方法往往需要大量特例设计, 导致实现复杂度和计算成本迅速上升, 进而限制可迁移性与实用性。

3.3 第 3 代方法: 基于符号树、嵌入表示或字符串表示的全开放符号空间

第 3 代方法的关键转变是进一步摆脱封闭候选集及其规则化扩展, 转向在更开放的表达空间中生成

和搜索方程结构。开放形式表示主要包括 3 种。第 1 种是符号树表示,即把方程表达为由变量、常数、算子和函数组成的树状结构,遗传编程、符号回归、树搜索和强化学习等方法可在这一空间中生成候选表达式。第 2 种是嵌入表示,即不直接在显式符号结构上搜索,而是将方程或其生成规则映射到连续隐空间向量,使搜索过程可以借助梯度优化、生成模型或相似性检索推进,并且更容易与神经网络端到端耦合;但这类方法通常需要额外的解码步骤,将隐空间结果还原为人类可读的符号表达^[8,27]。第 3 种是字符串表示,即把方程、代码或自然语言提示写成序列,使其像一种可读写的语言,从而可以借助大语言模型(large language model, LLM)在序列空间中生成、改写候选表达,并利用其内嵌的数学与科学知识作为启发式先验,在反馈驱动的迭代中逐步逼近可用方程^[28]。与封闭候选集和可扩展候选集相比,开放形式表示具有更强表达能力,但也更依赖严格的语法校验、物理约束、执行反馈与科学验证,以防止产生形式上合理但科学上无效的表达。

3.4 LLM 在知识发现中的多重角色及其风险

LLM 辅助科学发现也可以放在 RSE 框架中理解,其作用至少包括 5 类。第一,可基于预训练中获得的数学表达模式,生成候选方程,并借助数据评分迭代修正。但其输出可能存在语法不合法、变量定义不清、量纲不一致或违反物理约束等问题^[29]。第二,可生成代码并辅助完成计算实验^[30]。但需要防范代码错误、依赖缺失^[31]、环境不可复现和结果链条不可

追踪等风险^[32]。第三,可辅助文献检索与知识组织,但其输出可能存在引用幻觉、来源混淆和过度概括。例如, Walters 等^[33]发现, ChatGPT 生成的 636 条参考文献中, GPT-3.5 有 55% 完全虚构, GPT-4 的虚构比例也达 18%。Peters 等^[34]也指出, LLM 生成的科学摘要更容易作出超出原文证据范围的概括。第四,可生成实验设计,但可能忽略设备能力、成本限制、操作可行性和安全要求。第五,可为候选方程、模型行为和数据模式提供机制性解释。但这类解释仍可能存在因果误读、机制过度阐释等问题。总体而言, LLM 的局限包括表达式语法错误、变量定义不清、量纲不一致、物理约束缺失、引用幻觉、代码不可复现和结果解释过度等。因此, LLM 更适合作为开放表示空间中的候选生成器、交互接口和科研代理,而不能轻易替代物理一致性检验、实验验证和人类机制判断。

4 AI4S 知识发现的人机协作机制

AI4S 知识发现研究的发展方向是推动科学研究从传统的人类中心主义认识论逐步迈向人机协作范式^[35]。传统认识论默认以人类为中心,即认识目的由人设立、认识过程由人把握、认识结果由人裁决。进入数智时代后, AI 不再只是外部辅助工具,它能够在科学研究中分担人类认识活动,成为新的参与主体。如图 4 所示, AI4S 知识发现的人机协作机制是由碳基智能(人类)与硅基智能(机器)在认识差异所共同

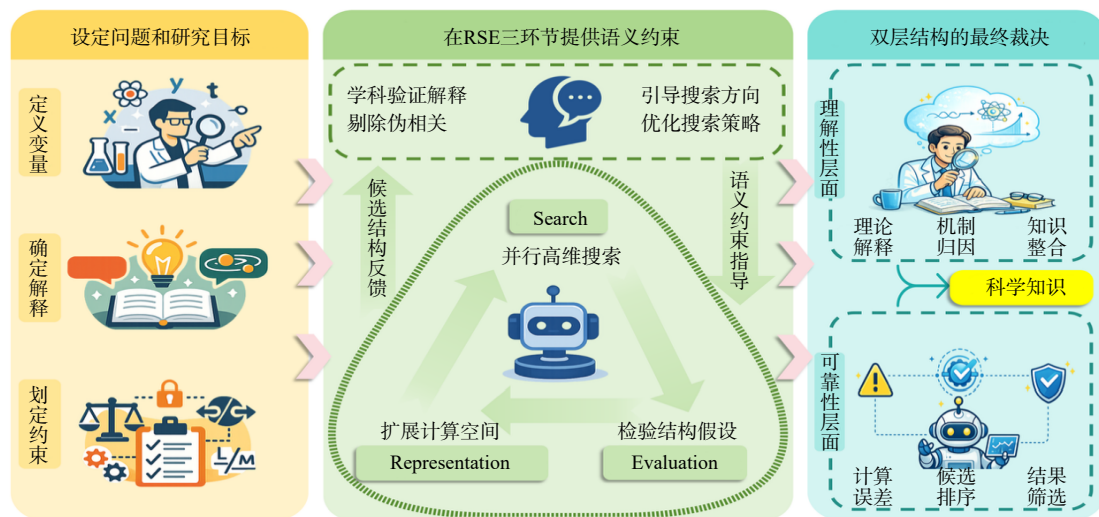


图 4 AI4S 知识发现的人机协作机制

决定的,机器更擅长对事实进行形式化处理,能够在高维强非线性数据中进行大规模搜索与拟合,生成候选方程或经验关系;而人类的认识不仅依赖逻辑推演,还包含非理性的直觉、灵感等因素^[36]。为理解人类所承担的角色,可拆解出3个认识任务:设定研究问题与认识目标、在RSE三环节提供语义约束、作出双层结构的最终裁决。

4.1 认知起点:人类界定问题与AI执行求解

对于设定研究问题与认识目标,人类科学家是不可替代的。正如普罗泰戈拉所言,“人是万物的尺度”。在科学发现中,这一尺度具体表现为决定究竟要认识什么、以何种语言表述并接受何种形式的解释与约束。这些问题并非简单的技术问题,而是涉及变量语义、物理量选择、允许的算子集合与组合规则、复杂度边界与约束形式(守恒律、对称性/不变性、量纲一致性等)等重要决策。这些决策高度依赖于领域知识、物理直觉与理论偏好,本质上属于一种意向性活动,即“为了何种目的去认识”的能力。因此,从固定候选项库的第1代方法到全开放符号表达空间的第3代方法,都是人类定义了问题,AI才去寻找答案。

4.2 发现过程:AI计算探索与人类语义约束

在RSE三环节中,AI在知识发现中承担的是可扩展的计算归纳与结构探索任务。随着方程表示空间从封闭走向开放,候选结构数量会出现指数级增长,并伴随高维非线性,远远超出人类工作记忆与串行推演的生物极限。稀疏回归、遗传/进化算法、强化学习,以及LLM辅助生成等方法,能够以可扩展的计算成本,在可接受的时间内对大规模结构空间进行寻优,压缩候选集并产出可供检验的结构假设。换言之,AI在这一范式中的优势是信息处理能力,即使不理解方程的物理含义,也能在数学空间中发现人类直觉难以触及的结构关联,从而形成对人类推理能力的延伸。

但这并不意味着把发现过程彻底外包给AI。首先,AI的信息处理能力并非无限通用,它的有效性仍受制于算法可形式化程度、搜索策略的可控性,以及复杂度可承受范围^[37]。其次,在语义约束缺失的情况下,AI往往依据观测数据拟合与预测表现来推理,可能学到的是背景相关,不一定对应真正关心的对象机制^[38]。即便给出事后解释(post-hoc explanation),也

只是经验性的补救说明,并不等价于揭示机制的因果解释^[39]。简言之,要让AI4S知识发现更可能逼近科学定律,必须把语义约束纳入RSE三环节。

4.3 结果整合:AI自动筛选与人类价值裁决

在RSE三环节后,人类科学家对候选集合进行解释与裁决,把结果纳入可解释的知识体系,并通过实验检验、理论推导或跨数据验证完成最终确认。这个价值裁决不同于RSE三环节内部的评估(如前文所述,包含精确性、简洁性、物理一致性、数学易解性4项判准)。因为算法即便给出了一组在判准意义下足够好的方程,却无法独立判断哪一个更可能与既有理论体系兼容,哪一个更符合机制解释,哪一个在跨情景迁移时更可信,哪一个更能生成可检验的新预测,并能引导后续实验或数值验证。在这些赋义、归因与整合的任务中,人类科学家的作用也是不可替代的。算法层面的定量指标可提供重要的可靠性证据,但这些证据主要基于统计相关性来支持可用性与稳健性;而科学知识的成立还需满足因果律与可解释性的要求^[40]。哪一个方程值得被视为科学知识的判断,本质上是一种面向科学知识体系的选择,依赖研究者对物理意义、理论结构与解释路径的理解。不仅如此,“人类主导的科学发现不仅依赖理性,还涉及顿悟、直觉和灵感等非理性因素的参与,这些正是人类在科学发现与发明中的优势所在。^[41]”综上,价值裁决呈现为一种双层结构,在可靠性层面,AI负责提供计算误差、稳定性与一致性指标,并据此对候选进行排序与筛选;在理解性层面,人类科学家承担理论解释、机制归因与知识整合。只有经由这一终审环节,机器产出的数学表达式才可能从拟合良好的模型,进入可解释、可检验、可迁移的科学理论体系。

综上所述,AI4S的方法论指向并非以人工智能替代人类科学家,而是构建了一种连接经验归纳与理论演绎的混合认识主体系统。人类提供目标设定、语义约束与价值裁决,人工智能在搜索环节承担高维结构探索与计算归纳任务。随着方程表示空间的持续开放,搜索复杂度不断上升,对计算智能的依赖将进一步增强。这并未削弱人类的作用,反而对科学家表示约束与评估判断的要求也会同步提高。因为AI4S的人机协作不仅是静态的任务分工,更是人与机器在认识活动中交互融合、共同进化的新型实践。

5 AI4S 知识发现对科研生态的影响

如果说 AI4S 知识发现首先改变的是科学发现如何发生,那么它接下来触及的便是谁在发现、以何种方式发现。如图 5 所示,这种影响贯穿 3 个层面:在微观层面,人机协作重塑了科学家的能力结构;在中观层面,组内与组间科研协作模式随之改变;在宏观层面,则催生出新学术身份认同。不过在展开讨论前,有一个前提需要先行澄清。AI4S 知识发现对科研生态的重塑并非在所有学科均匀铺开,而更可能集中于那些可形式化、可自动化、数据密集且验证成本较低的领域。可形式化是指研究对象能被写成可计

算的表示,如分子图、晶体结构或控制方程,这是归纳自动化得以介入的前提;可自动化是指存在机器人或程序化接口,能把 AI 的输出直接转化为实验动作;数据密集是指存在大规模、机器可读且标注一致的语料;验证成本低则意味着一个候选结果能被快速而明确地检验。计算化学、无机材料、方程发现等领域恰好在这 4 个维度上都得分较高,因而走得最快。相反,对于那些高度依赖现场经验、数据密度低、实验难以标准化或伦理约束较强的领域,问题定义、机制判断与价值取舍仍主要掌握在人类研究者手中。正因如此,下文论及的角色、组织与协作变化都只是条件性趋势,将其推广到所有学科时仍须审慎。

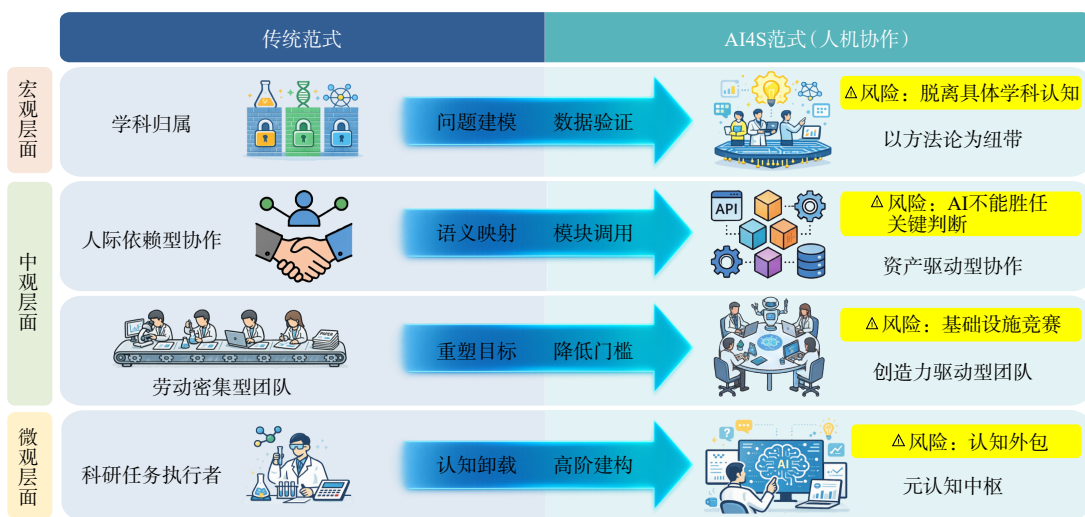


图 5 AI4S 知识发现对科研生态的影响

5.1 科学家角色从执行者到元认知中枢

AI4S 知识发现正在改变科学家的角色,使其从科研任务的执行者,转向更具元认知色彩的提问者、决策者、审核者和验证者。过去,科学共同体的分工主要围绕计算、实验与理论推演展开。而在 AI 逐步接管归纳、搜索与生成之后,人类的核心价值不再是解决问题,而是判断什么值得被问、哪些证据可信、结论是否成立,以及如何将结果纳入可解释的知识体系。

这一转变在实验科学中体现得尤为直接,因为 AI 正在接管其中反复试错的重复性操作。Boiko 等^[30]开发的 Coscientist 系统以 LLM 为中枢,能够(半)自主地设计、规划并执行化学实验,在钯催化交叉偶联等任务上完成了过去需多人分工的筛选与优化;Xu

等^[15]构建的自动化 TLC 平台,由 2 台协作机器人配合相机与控制软件,自主完成点样、展开、成像与 R_f 计算的全流程,生成了涵盖 387 种化合物、17 种洗脱条件的 4944 条标准化数据;Szymanski 等^[42]的自主实验室 A-Lab 则把机器人、第一性原理数据库、机器学习解释、文献挖掘与主动学习整合于一体,用于无机材料的固态合成,在 17 d 连续运行中针对 58 个目标合成出 41 种新化合物。值得注意的是,这 3 个案例都集中在化学与材料领域,而这恰恰印证了前文的判断——它们正是可形式化、可自动化、数据密集且验证成本低的领域。换言之,AI 作为执行者的角色是有边界的。在那些缺乏成熟自动化接口的领域,人工操作仍难以替代。即便在材料与化学等条件较为成熟的领域,研究者也强调 AI 自动化产出仍须独立验

证。可见, AI 接管的是执行, 而非判断。

实验科学的角色重构受制于物理操作与自动化接口, 但软件开发的角色重构似乎面临更少阻力。软件的产出本身即是代码, 不存在硬件执行瓶颈。因此, 前端开发、后端开发、算法开发、测试、部署等以往需要多人分工的工作, 在自动编程与智能代理介入后, 更容易被整合为个人即可驾驭的全栈化执行能力。但无论是实验科学还是软件开发, 科学家角色重构的关键不在于 AI 替代了多少劳动, 而在于它把科研的主战场从解题能力推向问题能力, 从技能型劳动转向判断型劳动, 从局部优化转向方向选择。

这种角色转变呈现出典型的双刃剑效应。积极的一面是认知卸载(cognitive offloading)。模式搜索与候选方案生成等工作被转移给 AI, 人类得以把有限的注意资源投入到更高阶的理论建构、机制推理与价值判断上。这不仅为科学直觉腾出了空间, 也直接提升了科研效率与产出密度。然而, 消极的一面是认知卸载极易滑向认知外包(cognitive outsourcing)。原本希望外包的是低阶重复劳动, 但如果在表示、搜索和评估这 3 个环节丢失人类作为认识主体的职责, 那么被外包的可能是问题意识、方向选择、因果直觉、价值权衡与学术品味^[43]。科研看似更快, 实则可能在更深层面变得更空心化, 即产出密度上升, 知识密度下降。

5.2 科研团队结构组内更扁平、组际更分化

随着 AI4S 知识发现提升单人的全栈化执行能力, 并将科学家角色推向元认知中枢, 它也可能相应地改变科研组织的方式与培养目标。在小型科研团队内部(如独立项目负责人带领的课题组), AI4S 知识发现或将改写科研中长期存在争议的劳动密集型“流水线”团队结构。以化学、生物医学等实验学科为例, 为提高产出效率, 表征、合成、计算、绘图写作往往被拆解为可并行的岗位, 年轻的科研工作者得以快速习得相关技能并产出成果, 代价却是牺牲了本应锻炼的整体科研能力。AI4S 知识发现恰好可能对此类模式形成替代: 那些高度标准化、重复性强、以规则化操作为主的环节(数据清洗与标注、常规拟合与对照实验设计、文献检索与信息抽取、图表生成与版本化写作等), 相对更容易被 AI 承接并降本增效。如果这一替代机制成立, 能够留存并发展的课题组, 便有条件

把人力更多聚焦于问题定义、假说建构、证据审查、机制解释与知识整合, 而非低阶重复性工作——这或许有助于科研人才培养回归本质。

与此同时, AI 降低了工具门槛, 可能缓解技能差异造成的机会不均, 使更多研究者获得把想法转化为研究方案的能力, 从而在一定程度上推动科研平权化^[44]。但平权化未必走向一个更加公平的学术共同体, 也可能伴随结构性不平等的加剧。由于算力、数据、模型、工具链和工程化能力具有强烈的资源属性, 容易形成“强者更强”的马太效应(Matthew effect)。在此条件下, AI4S 未必能自动拉平差距, 反而可能把科研竞争推向基础设施竞赛, 使知识生产进一步向少数高资源节点集中。平权与集中这 2 种倾向孰强孰弱, 很大程度上取决于治理, 并非技术本身的既定结局。

综上, AI4S 对课题组内部的改变, 可能是一场更深层的角色与结构再分配, 它有望压缩低阶重复劳动的需求, 倒逼课题组回归培养科学家的本质目标, 但同时也可能把竞争推向算力、数据等基础设施竞赛, 放大组际分化与累积优势。而这种组际分化, 对占据资源高地的优势课题组而言, 又意味着新的战略选择: 是选择“赢者通吃”以维持优势, 还是选择开放共享, 通过平台化供给实现更大范围的成果放大与价值外溢? 从现实约束来看, AI4S 的高效运行往往高度依赖多维度的数据输入与跨学科的交叉验证, 单纯的封闭垄断未必可持续^[45]。因此, 在一定条件下, 组际分化也可能转化为跨课题组、跨学科协作的起点, 推动形成资源互补、协同创新的科研格局。

5.3 科研协作机制从人际依赖到资产驱动

当研究推进速度更快、试错成本更低、跨域机会更密集时, 跨课题组协作将更频繁地被触发。在跨课题组协作层面, AI4S 未必能完全打破学科壁垒, 但它有望以更务实的方式降低协作成本, 并在一定程度上重塑协作机制与治理逻辑。

首先, 跨学科合作的核心痛点之一在于如何将不同学科的概念映射到共同的对话池中并流动起来, 这对人类科学家的“语言能力”提出了很高要求。一个直观案例是, 在计算机视觉里把低分辨率的数据场恢复成高分辨率通常被称为超分辨率(super-resolution), 而在地学与气候科学里, 非常相似的目标往往

称作降尺度(downscaling)。两者背后的数学任务高度重叠,本质都是在多尺度信息约束下进行从粗到细的重建与推断。但如果停留在术语层面,很容易误以为是2套互不相干的方法体系。AI在这里可以充当翻译器,把原本用于理解对方学科背景的沟通成本,压缩为定义问题并跑通流程的对齐问题——输入是什么、输出是什么、约束是什么、评价指标是什么、失败边界是什么。从这个意义上看,AI4S知识发现像是一座跨学科的“巴别塔”反向工程,它不消灭差异,但可能让差异可映射、可对齐、可检验。

其次,一个在条件成熟时可能率先出现的趋势是,AI4S知识发现把跨组协作从关系型推向资产驱动。过去合作往往取决于认识谁、对方愿意投入多少;而AI4S知识发现可能让协作越来越像调用资产。例如,一个掌握流体力学方程发现方法的团队,不必深度介入医学知识,也可以先贡献部分模块,帮助医学影像团队快速完成初步建模与验证。交叉学科研究便有可能从兴趣驱动转向收益激励下的理性选择,因为一份劳动可沉淀为可复用的资产,通过协作发挥杠杆效应、提高边际收益。这类资产化协作已有现实载体。材料计算平台Materials Project以开放方式提供大规模计算所得的材料信息与分析工具,并通过开源库pymatgen与公开应用程序编程接口(application programming interface, API)供外部团队直接调用^[46]。前述A-Lab的目标化合物,正是借助该平台的从头算相稳定性数据筛选得到的^[42]。这说明一个团队沉淀的数据与方法,确实可以作为“可复用资产”被另一团队在不深入其学科细节的前提下调用,从而降低协作的前置成本。但模块化与资产化并不意味着脱离人际协作而自动流转。大量关键环节仍属于隐性知识,如哪些异常值得当作信号、现象是否可靠、边界条件是否成立,往往依赖长期浸润与现场经验。因此,AI4S知识发现或许能让前期对齐与推进更快,但依然需要人际协作完成关键判断与验证。

最后,随着AI更深地介入科研生产环节,跨学科协作的信任结构也可能从共识驱动转向验证驱动。因为最大的风险并非错误本身,而是错误更像正确。这意味着跨组合作或将越来越难以依靠经验权威或直觉共识维持,而更多需要可复现实验、统一的基准

测试、版本管理与溯源记录。这一信任结构的转变已有初步实证。一项面向100位化学专家的调查显示,相较传统黑箱模型,可解释模型获得了更高的专家信任,且这一倾向在对AI理解更深的化学家中更为明显^[9]。这初步表明,当AI的产出从不可检视的预测转为可被审查的显式规律时,专家信任更容易建立——即信任的基础正从经验权威转向对结果本身的可验证性。

综上所述,AI4S尚未让学科壁垒消失,但它可以降低跨学科协作的沟通成本,不必完整理解对方的学科背景,而是先完成接口层面的对齐,再围绕可检验的任务进行分工与验证。在此条件下,协作模式可能从以人脉关系为核心的关系型合作,逐步转向以数据、流程与工具链为核心的资产化调用。跨组协作的信任机制也可能从依赖经验权威与直觉共识,转向以可复现流程、基准评测与溯源记录为基础的验证驱动。需要强调的是,这些是在数据密度、自动化与验证条件较成熟时可能率先显现的趋势,而非已普遍发生的事实。

5.4 AI4S知识发现的学术共同体认同

人机协作与人际协作的变化最终将触及更深层的结构,AI4S知识发现研究者的学科身份认同正在发生微妙转变。AI4S知识发现研究者可能渐渐地较难仅凭单一学科来定义,而更多被识别为能把科学问题转化为可计算对象的人,从学科归属转向方法归属,并可能促成一个以方法论为纽带的学术共同体。与之相匹配,其至少需要具备3类关键能力:其一是问题建模能力,即把现象拆解为变量、约束与可验证假设;其二是工具链装配能力,即选择合适表征、数据策略与模型路径并形成可运行的方案;其三是解释—验证闭环能力,即将模型输出重新嵌回科学解释与实验检验,使结论能够经得起证据审查与机制追问。

随着一个以方法论为纽带的学术共同体初现雏形,对AI4S知识发现研究者的核心能力边界也需要重新审视。如果方法论的学科身份被过度强调,看似在生物、物理、化学等多个领域“无所不能”(AI for everything),可迁移性极强,实则缺乏学科深水区的支撑。事实上,隐性知识与解释权仍掌握在真正扎根领域的人手里,只有这些人能把结果拉回科学真实。因此,AI4S知识发现并不是用方法“替代”学科,而更像

用方法“连接”学科。为此, AI4S 知识发现研究者既要与学科专家形成互补, 同时也要对自身能力边界有清醒的觉察——其方法论的可迁移性, 是否深入到科学表征与机制约束层面, 还是仅停留于模型架构的复用与性能指标的迁移。唯有形成优势互补与自我觉察, 方法论才能避免沦为一种无根的游历, 真正扎根于具体学科的认知土壤, 并在连接不同学科的间隙中生长出新的理论果实。

6 结论

基于上述分析, 建议通过接口对齐推动跨团队分工, 通过资源调用放大方法与数据的协同效应, 再以验证闭环守住科学研究的基本底线, 并以“方法连接学科”为导向培养“元认知中枢”型科研人才。当 AI 逐步接管归纳、搜索与生成, 人类科学家的职责将更集中地体现在提问、决策、审核与验证中。目标是以复现立信, 构建可信赖的研究基础; 以解释立论, 深化对科学现象的理解; 以迁移立功, 拓展知识的应用边界。这样, 科研范式的升级才真正转化为可累积的知识增量, 持续推进人类认知前沿发展。

参考文献 (References)

- [1] The White House. Launching the genesis mission[EB/OL]. (2025-11-24) [2025-11-25]. <https://www.whitehouse.gov/presidential-actions/2025/11/launching-the-genesis-mission/>.
- [2] Chen Y T, Zhang D X. Theory-guided deep-learning for electrical load forecasting (TgDLF) via ensemble long short-term memory[J]. *Advances in Applied Energy*, 2021, 1: 100004.
- [3] Zhang D S, Chen Y T, Chen S Y. Filtered partial differential equations: A robust surrogate constraint in physics-informed deep learning framework[J]. *Journal of Fluid Mechanics*, 2024, 999: A40.
- [4] Vass M, Kooistra A J, Ritschel T, et al. Molecular interaction fingerprint approaches for GPCR drug discovery[J]. *Current Opinion in Pharmacology*, 2016, 30: 59-68.
- [5] Chen Y, Zhang D. Integration of knowledge and data in machine learning[EB/OL]. (2022-02-15) [2025-11-25]. <https://arxiv.org/abs/2202.10337>.
- [6] 江俊, 崔乘幸, 黄文光. AI能否成为化学知识的发现者?[J]. *科技导报*, 2025, 43(21): 16-22.
- [7] 麦克莱伦, 多恩. 世界科学技术通史[M]. 王鸣阳, 译. 上海: 上海科技教育出版社, 2007.
- [8] Chen Y T, Luo Y T, Liu Q, et al. Symbolic genetic algorithm for discovering open-form partial differential equations (SGA-PDE)[J]. *Physical Review Research*, 2022, 4(2): 023174.
- [9] Lou S Y, Liu C C, Zhang D X, et al. Unsupervised hierarchical symbolic regression for interpretable property modeling in complex multi-variable systems[J]. *Advanced Science*, 2026, 13(19): e21200.
- [10] Wang D, Feng D C, Zhou K C, et al. Symbolic regression-enhanced dynamic wake meandering: Fast and physically consistent wind turbine wake modelling[J]. *Journal of Fluid Mechanics*, 2025, 1025: A56.
- [11] Zeng J S, Xu H, Chen Y T, et al. Deep learning discovery of macroscopic governing equations for viscous gravity currents from microscopic simulation data[J]. *Computational Geosciences*, 2023, 27(6): 987-1000.
- [12] Xu H, Chen Y T, Zeng Z Z, et al. Exploring terrain-precipitation relationships with interpretable AI for advancing future climate projections[J]. *Nexus*, 2025, 2(1): 100045.
- [13] Ross A, Li Z W, Perezogin P, et al. Benchmarking of machine learning ocean subgrid parameterizations in an idealized model[J]. *Journal of Advances in Modeling Earth Systems*, 2023, 15(1): e2022MS003258.
- [14] 王东. 人工智能与科学发现: 一种哲学探究[M]. 北京: 知识产权出版社, 2023.
- [15] Xu H, Lin J L, Liu Q Y, et al. High-throughput discovery of chemical structure-polarity relationships combining automation and machine-learning techniques[J]. *Chem*, 2022, 8(12): 3202-3214.
- [16] Chen Z, Liu Y, Sun H. Physics-informed learning of governing equations from scarce data[J]. *Nature Communications*, 2021, 12: 6136.
- [17] Both G J, Choudhury S, Sens P, et al. DeepMoD: Deep learning for model discovery in noisy data[J]. *Journal of Computational Physics*, 2021, 428: 109985.
- [18] Xu H, Chang H B, Zhang D X. DL-PDE: Deep-learning based data-driven discovery of partial differential equations from discrete and noisy data[J]. *Communications in Computational Physics*, 2025, 29(3): 698-728.
- [19] Wang D, Chen Y T, Chen S Y. Discovering an interpretable mathematical expression for a full wind-turbine wake with artificial intelligence enhanced symbolic regression[J]. *Physics of Fluids*, 2024, 36(10): 105110.
- [20] Takaki K, Miyao T. Symbolic regression for the interpretation of quantitative structure-property relationships[J]. *Arti-*

- ificial Intelligence in the Life Sciences, 2022, 2: 100046.
- [21] Brunton S L, Proctor J L, Kutz J N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113(15): 3932–3937.
- [22] Xu H, Zhang D X, Wang N Z. Deep-learning based discovery of partial differential equations in integral form from sparse and noisy data[J]. *Journal of Computational Physics*, 2021, 445: 110592.
- [23] Kaheman K, Kutz J N, Brunton S L. SINDy-PI: A robust algorithm for parallel implicit sparse identification of nonlinear dynamics[J]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2020, 476(2242): 20200279.
- [24] Niessner J, Helmig R, Jakobs H, et al. Interface condition and linearization schemes in the Newton iterations for two-phase flow in heterogeneous porous media[J]. *Advances in Water Resources*, 2005, 28(7): 671–687.
- [25] Kaptanoglu A, de Silva B, Fasel U, et al. PySINDy: A comprehensive Python package for robust sparse system identification[J]. *Journal of Open Source Software*, 2022, 7(69): 3994.
- [26] Xu H, Chang H B, Zhang D X. DLGA-PDE: Discovery of PDEs with incomplete candidate library via combination of deep learning and genetic algorithm[J]. *Journal of Computational Physics*, 2020, 418: 109584.
- [27] Du M G, Chen Y T, Zhang D X. DISCOVER: Deep identification of symbolically concise open-form partial differential equations via enhanced reinforcement learning[J]. *Physical Review Research*, 2024, 6: 013182.
- [28] Du M G, Chen Y T, Wang Z Z, et al. Large language models for automatic equation discovery of nonlinear dynamics[J]. *Physics of Fluids*, 2024, 36(9): 097121.
- [29] Shojaee P, Meidani K, Gupta S, et al. LLM-SR: Scientific equation discovery via programming with large language models[J/OL]. arXiv, 2024: 2404.18400. <https://arxiv.org/abs/2404.18400>.
- [30] Boiko D A, MacKnight R, Kline B, et al. Autonomous chemical research with large language models[J]. *Nature*, 2023, 624(7992): 570–578.
- [31] Tambon F, Moradi-Dakhel A, Nikanjam A, et al. Bugs in large language models generated code: An empirical study[J]. *Empirical Software Engineering*, 2025, 30(3): 65.
- [32] Sandve G K, Nekrutenko A, Taylor J, et al. Ten simple rules for reproducible computational research[J]. *PLoS Computational Biology*, 2013, 9(10): e1003285.
- [33] Walters W H, Wilder E I. Fabrication and errors in the bibliographic citations generated by ChatGPT[J]. *Scientific Reports*, 2023, 13: 14045.
- [34] Peters U, Chin-Yee B. Generalization bias in large language model summarization of scientific research[J]. *Royal Society Open Science*, 2025, 12(4): 241776.
- [35] 保罗·汉弗莱斯. 延长的万物之尺: 计算科学、经验主义与科学方法[M]. 苏湛, 董春雨, 孙卫民, 译. 北京: 人民出版社, 2017.
- [36] 董春雨, 王逸凡. 数智时代有关机器认识主体性问题的若干探讨[J]. 当代中国价值观研究, 2023, 8(6): 17–28.
- [37] 董春雨. 从机器认识的不透明性看人工智能的本质及其限度[J]. 中国社会科学, 2023(5): 148–166, 207, 208.
- [38] 贾玮晗, 董春雨. 论数据模型的特征及其认知功能[J]. *自然辩证法通讯*, 2024, 46(4): 23–30.
- [39] 贾玮晗, 董春雨. 事后解释可以消除认识的不透明性吗[J]. *探索与争鸣*, 2024(6): 80–87.
- [40] 董春雨, 郭艳娜. 认识黑箱视角下相关性与因果性关系之辨析[J]. *自然辩证法研究*, 2020, 36(12): 54–59.
- [41] 贾玮晗, 董春雨. 机器可以实现科学发现吗: 机器智能在科学发现中的价值与限度[J]. 北京师范大学学报(社会科学版), 2025(1): 151–158.
- [42] Szymanski N J, Rendy B, Fei Y X, et al. An autonomous laboratory for the accelerated synthesis of inorganic materials[J]. *Nature*, 2023, 624(7990): 86–91.
- [43] 王从余, 官平. 人工智能与人类共同创造的范式演进与实践路径[J]. *科普研究*, 2025, 20(5): 5–16, 47, 123.
- [44] Eapen T, Finkenstadt D J, Folk J, et al. How generative AI can augment human creativity[J]. *SSRN Electronic Journal*, 2024: 56–64.
- [45] 匡光力, 黄杏洁, 汪文强. 建好、用好大科学装置积极推动人类知识创新: 以稳态强磁场实验装置为例[J]. 科技导报, 2025, 43(5): 26–36.
- [46] Jain A, Ong S P, Hautier G, et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation[J]. *APL Materials*, 2013, 1: 011002.

Paradigm transformation and ecological reconstruction of AI-driven scientific discovery

WANG Congyu¹, MI Haoyuan¹, QI Wenxiu², CHENG Peng³, ZHANG Dongxiao^{4*}

1. College of Investigation, People's Public Security University of China, Beijing 100038, China

2. Department of Philosophy, Peking University, Beijing 100871, China

3. China Research Institute for Science Popularization, Beijing 100081, China

4. Eastern Institute of Technology, Ningbo, Ningbo 315200, China

Abstract AI for Science (AI4S) is reshaping the paradigm of scientific inquiry by integrating scientific priors with data-driven methods. Within this broader shift, AI-driven scientific discovery has become a question of national strategic importance. This paper examines its conceptual boundaries, methodological landscape, and human-AI collaboration mechanisms, and considers its potential implications for the research ecosystem. We first distinguish the two complementary pathways of AI4S—knowledge embedding and knowledge discovery—and the closed loop they form, and argue that the core contributions of knowledge discovery lie in automating inductive reasoning and explicit knowledge formalization. This contribution, however, is concentrated at the level of phenomena and empirical laws; the leap to constructive theories still depends on human mechanistic interpretation and validation. We then propose the Representation–Search–Evaluation (RSE) framework, showing that knowledge–discovery methods differ in the openness of their representation space, with search and evaluation co-evolving around that openness. On this basis, we characterize the core responsibilities that remain with humans in human–AI collaboration: defining epistemic goals, providing semantic constraints, and leading value-based adjudication. Finally, we discuss how knowledge discovery may reshape the research ecosystem. In fields where data density, automation, and validation conditions are relatively mature, the role of scientists may shift from task executors toward a kind of “metacognitive hub” and research organizations may become flatter within groups while growing more differentiated across groups. Such differentiation can be turned into a higher-quality basis for cross-group collaboration through platformized assets, standardized interfaces, and effective validation governance. We emphasize that these ecosystem-level claims are conditional tendencies rather than established facts. The paper’s innovation is to articulate, within the RSE framework, both the methodological trajectory of knowledge discovery and its associated human–AI collaboration mechanism, and on that basis to offer an outlook on the possible changes in research collaboration and organization.

Keywords artificial intelligence; AI for Science; knowledge discovery; human–AI collaboration ●



(责任编辑 王微)