Insight Note | 智辨系列

总 No.5 期/2026 年 9 月

AI 时代科学研究范式的演进:
基于复杂系统与衍生论的跨学科综述

序言

人类对自然规律与宇宙秩序的探索，在“还原”（Reductionism）与“涌现”（Emergence）这两种核心认识论之间交替演进。在探讨宇宙初始之物质根基（东方哲学所谓之“太素”）与万物运行之客观数理法则（西方哲学所谓之“逻各斯”）的过程中，科学界逐渐认识到，单一的向下解构已无法完全解释宏观世界的复杂性。

随着人工智能（AI）技术的爆发式发展，特别是大参数模型在海量数据处理与高维空间映射表现出的卓越能力，“AI for Science”（人工智能驱动的科学研究）正在引发一场深刻的范式革命。2026 年 6 月 30 日，上科大的“跨学科下午茶-太素 logos 系列”围绕这一主题进行了无界漫谈。本综述旨在以物理学的发展史为思想基底，深入剖析还原论的局限与衍生论的崛起，并系统探讨人工智能在物质科学、生命科学、社会经济系统以及临床与工程实践中，如何作为一种跨尺度整合工具，重构现代科学的研究逻辑。

第一章 认识论的历史演进：从还原论的丰碑到衍生论的觉醒

张广铭教授的分享表明，科学史的演进，在很大程度上是人类对物质世界结构认知不断深化的历史。在长达数百年的时间里，还原论构成了自然科学最底层的信念体系。还原论的核心假设在于：一切复杂的自然现象，均可严格归结为构成该系统的最基本结构单元，以及决定这些微观单元行为的最基本物理规律。这种哲学的最终目标是建立一个包罗万象的大统一理论（Theory of Everything），假定只要掌握了底层的“积木”和支配它们的“法则”，人类就能自下而上地推演并重建整个宇宙。

1.1 物理学的四次科学革命与还原论的巅峰

物理学作为最基础的成熟科学，其历史经历了四次公认的重大科学革命。每一次革命都极大地统一了先前的认知分裂，并带来了全新的世界观与数学工具，将还原论的成就推向顶峰。

首先是 17 世纪由牛顿发起的力学革命。牛顿力学统一了天体运动规律与地球上物体的运动规律，确立了“物质由粒子构成”的世界观。在这个体系中，微观粒子的运动遵循确定性的牛顿运动定律，系统时间是完全可逆的，任何物理过程理论上都可以反演复现。微积分作为其核心数学工具，为描述这种绝对确定性提供了完美的语言体系。

进入 19 世纪，麦克斯韦主导了电磁革命，完成了电、磁与光的历史性统一。这一革命揭示了物质世界除了离散的粒子实体外，还存在一种连续的物质形态——波。波被视为传递带电或磁性粒子间相互作用的介质形态。规范场理论的引入，使得人类能够精确描述涵盖微波、红外线、可见光、紫外线、X 射线直至伽马射线的广阔电磁波谱，极大地扩展了人类对物质形态丰度的认知边界。

20 世纪初，爱因斯坦的相对论革命统一了空间、时间与引力。相对论彻底摒弃了牛顿体系中的绝对时空观与“以太”假说，指出时空本身

具有动力学属性，可以发生畸变、弯曲，并以引力波的形式传递时空扰动的能量。黎曼几何（弯曲空间的几何学）成为描述这一宏大宇宙图景的基础数学工具，确立了引力场方程的普遍有效性。

几乎在同一历史时期，由多位杰出物理学家共同缔造的量子革命，实现了经典粒子与经典波的深层次统一，提出了震撼科学界的“波粒二象性”。微观世界的物质既不是纯粹的经典粒子，也不是纯粹的连续波，而是一种遵循线性代数与张量分析规律的量子态。在学术界，有学者常常使用视觉图像学中的错觉图形（如同一张图片在不同视觉聚焦点下呈现为少女的侧面或老妇人的正面）来隐喻量子行为：量子体系的客观观测结果高度依赖于观测者的前提条件、视角与特征边缘的信息整合方式，但它们本质上是同一客观存在的不同表象。这一特征标志着人类对微观物质世界基本规律的认知达到了前所未有的深度。

科学革命阶段	核心代表人物/时代	统一的物理领域	带来的全新世界观	核心数学工具
力学革命	17 世纪 (牛顿)	天体运动与地球物体运动	粒子观、确定性、时间可逆	微积分
电磁革命	19 世纪 (麦克斯韦)	电、磁、光	波动观、相互作用介质	规范场理论
相对论革命	20 世纪初 (爱因斯坦)	时间、空间、引力	动态弯曲时空、引力波	黎曼几何

量子革命	20 世纪初（多位物理学家）	粒子属性与波动属性	波粒二象性、量子叠加态	线性代数、张量分析

在上述革命的推动下，还原论指导着物理学家将物质结构的层次层层剥开，建立起了一套极为严密的特征尺度体系。

物质结构层次	特征空间尺度	特征能量尺度
生物大分子	10^{-5} 至 10^{-7} 厘米	0.1 至 几个电子伏特
原子	10^{-8} 厘米	几个电子伏特
原子核	10^{-13} 厘米	兆电子伏特（10^{6} eV）
质子/中子	10^{-13} 厘米量级	十亿电子伏特（10^{9} eV）
夸克	10^{-16} 厘米	极高能标

伴随空间特征尺度从大分子向夸克的不断缩小，特征能量呈指数级跃升。至此，物质结构的最底层物理规律似乎已被完全穷尽。

1.2 还原论的局限与热力学挑战

然而，科学界在 20 世纪末逐渐达成共识：将万事万物还原为最基本的物理规律，并不蕴含着从这些微观规律出发重建整个宏观复杂系统

的能力。1996 年，科普作家约翰·霍根（John Horgan）在其引发广泛争议的著作《科学的终结》中尖锐地指出，如果纯粹基于还原论的思想，科学探索之路已面临认识论上的极限，依靠底层粒子的运动方程根本无法推导出蛋白质的高级折叠功能或人类大脑的意识行为。

宏观自然现象中普遍存在的“不可逆性”构成了对经典力学还原论的第一次重大挑战。以一滴红墨水在清水中的扩散过程为例，该宏观演化是单向且不可逆的，墨水一旦扩散便无法自发重新聚集成一滴，这完全违背了牛顿力学中时间绝对可逆的确定性原则。为了解释这种宏观过程的不可逆现象，热力学这一全新学科分支应运而生。

德国科学家克劳修斯（Clausius）指出，在能量守恒（热力学第一定律）之外，自然界存在另一个同等基本的核心物理量——熵（Entropy）。对于一个没有能量交换的孤立系统，其动力学演化的铁律是始终朝着“熵最大”的方向进行。随后，奥地利物理学家玻尔兹曼（Ludwig Boltzmann）从微观统计学的角度为熵赋予了深刻的概率论内涵：熵是一个系统宏观状态出现可能性的测度（A measure of the probability of macrostate）。其数学定义极为简洁，即微观状态数取对数（log）后乘以玻尔兹曼常数。

这意味着，热力学第二定律所揭示的自然界演化方向，本质上是从小概率状态（如任意设定的高度有序初始态）向大概率状态（最可几的无序宏观态）演化。以抛掷 100 枚硬币为例，全部正面向上的构型（极度有序）仅有 1 种可能，概率极低；而正反各半的构型出现可能性最大，具有最大的熵值。这种“统计规律”的引入，打破了牛顿式的严格因果确定性，确立了概率与信息密度在宇宙演化中的核心地位。有学者指出，在系统复杂化的演化链条中，信息本质上即为“负熵”，系统的结构复杂化意味着其存储和传递的信息密度不断升高。

1.3 对称性破缺与衍生论（涌现论）的确立

在探讨物理系统时，不仅需要研究孤立系统，还需要研究允许能量

交换但无物质交换的“封闭系统”。对于封闭系统而言，其热力学演化过程由“自由能”（Free Energy，$F = U - TS$）主导。热力学第二定律指出，封闭系统最终达到的宏观平衡状态，必须是自由能极小的状态。然而，公式内部存在深刻的竞争关系：内能（U）的极小化驱使系统走向极度“有序”（如绝对零度下的理想晶体）；而熵（S）的极大化则驱使系统走向极度“无序”（如无相互作用力的理想气体）。这两股矛盾趋势在可调参数温度（T）的调节下相互妥协。当温度改变导致主导权发生转换时，系统便会经历从无序态到有序态，或从一种宏观态到另一种宏观态的激变，即所谓“相变”（Phase Transition）。

凝聚态物理学宗师朗道（Landau）凭借对相变的深刻洞见指出，自然界极其丰富的物质形态来源于不同的“序”（Order），而这些不同的序直接源于“对称性破缺”（Symmetry breaking）。以最常见的水为例，水分子（H_2O）在液态、气态、固态下的微观化学结构完全一致。一滴水中包含约 10^{23} 数量级的水分子，正是当数量级达到宏观阈值时，分子间产生了一种复杂的“集体运动”与“集体效应”，从而随着温度变化涌现出液态水、水蒸气以及拥有近 300 种不同结晶结构的冰。在三维空间体系中，破坏空间平移对称性的方式共有严格的 230 种，这 230 种微观结构单元的平移对称性破缺，构成了 230 种空间群的晶体，展现出包括铁磁性、超导性在内的极其丰富的宏观物性，成为 20 世纪中叶发展起来的固体物理学（Solid State Physics）的基石。

基于对称性破缺与相变理论，著名物理学家 P.W. Anderson 在 1972 年《Science》上发表了具有划时代意义的文献《多者异也》（More is Different），正式确立了“衍生论”（Emergence，亦称涌现论）的哲学地位。衍生论强调，客观物质世界是严格分层次的。在从微观基本粒子（夸克域）跨越物理学、化学、生物学，直至涌现出宏观社会系统与智能的阶梯式层级中，每一个层次都拥有极其专属的特征空间尺度、能量尺度和基本规律。

Anderson 的核心论断在于：面对尺度与复杂性的双重困难，还原论试图从极少数基本粒子的简单外推来解释复杂聚集体行为的假定崩塌。在复杂性的每一个新层次，都会涌现出全新的物理性质和“组织原理”（Organizing principle）。更为关键的是，层次与层次之间存在着明显的“脱耦”（Decoupling）现象——下一层次的微观现象对上一层次的宏观规律未必具有决定性的因果关系。正如中国哲学中的“整体论”思想所言，“整体远大于部分之和”。从宇宙大爆炸后的暴胀与星系螺旋结构的引力成形，到早期无机物向含碳有机分子、大分子蛋白质的演化，自然史的本质就是一部物质运动集体行为的衍生史。这种层级之间的脱耦与界面鸿沟，意味着低维法则在高维系统面前必然失效，必须依赖统计力学、相变理论以及当代的人工智能大模型来建立新的高层映射逻辑。

第二章 物质科学的跨尺度鸿沟与 AI 计算重构

在确立了衍生论的现代科学观后，物质科学研究面临的最大痛点清晰浮现：如何弥合不同特征尺度之间的“交界面”（Interface）。在特定的极微观或极宏观尺度内，传统理论均已建立了逻辑自洽的严密闭环；但当研究涉及多尺度耦合时，理论断层便成为不可逾越的天堑。

2.1 理论断层与催化化学的多尺度挑战

以上海科技大学等科研机构高度关注的多相催化与化学反应工程为例，其研究在时空跨度上呈现出碎片化特征。在底层量子力学层次，研究聚焦于原子、电子级别的微观相互作用（皮秒级时间、埃米级空间）；在此之上，研究需过渡至极快的超快化学反应动力学，进而涉及固体催化材料本体结构在反应过程中的缓慢相态演化；最终，系统必须放大到涉及流体力学、宏观传递、扩散工程的工业级反应器尺度（小时级时间、米级空间）。

在这一层层递进的七八个层次之间，微观层面上（如纳米结构的些微修饰）的极小变量，一旦经过 10^{23} 次方量级的原子统计累积与非线性放大，往往会导致宏观催化现象发生天翻地覆的相变异动。传统的还原论计算手段（如第一性原理计算与密度泛函理论，DFT）虽然能够极度精准地求解极少数原子的相互作用，但当面对宏观统计行为时，计算量呈现指数级爆炸，完全无法实现微观变量对宏观最佳催化放大效果的精准外推预测。如何在交界面处将上一个层次的输出无损转换为下一个层次的输入，成为了现代物理化学界悬而未决的核心难题。

2.2 人工智能大模型的跨界关联与赋能

为了跨越这一理论断层，学界正在将视野转向数据驱动与人工智能。AI 的核心优势在于，它不依赖于解析公式强行跨界，而是通过海量数据训练，建立起高维度的模式识别（Pattern Recognition）与超

强相关性挖掘机制，从而有效填补理论断裂带来的鸿沟。

在这一前沿阵地，上科大学者的学术建树尤为值得关注。例如，物质学院副院长张广铭教授，其学术生涯长期致力于理论凝聚态物理、介观近藤屏蔽效应（Kondo screening effect）与二维拓扑绝缘体激发等领域的研究。其对量子物理体系与衍生论基本规律的宏观把控，为 AI 在复杂系统中的底层逻辑提供了深厚的理论指导框架。

在具体的化学计算模型落地方面，物质学院的胡培君教授团队取得了突破性进展。胡培君课题组成功发布了新一代通用原子大模型 TACE（Tensor-based Atomic representation via message-passing for Chemical systems and beyond）。在 Matbench Discovery 全球排行榜的严格评测中，TACE 模型展现出了极大的技术优势。相较于当前主流的大规模球谐模型（如 MACE, GRACE, Allegro 等），TACE 能够以更加精简的模型体量，实现甚至超越庞大模型的预测精度与复杂体系表达能力。更为核心的是，TACE 框架深度结合了团队自主研发的、专注于多相催化反应体系的 REICO 数据集构建策略（发表于《Nat Catal》）。这种“数据-模型”的双重底层创新，使得 AI 不再是黑盒，而是真正成为了催化剂理性设计与复杂反应机理精确解析的强有力计算引擎。

胡培君教授指出，未来的 10 年将是 AI 赋能科学研究的黄金发展期，科研工作者群体可能会经历一次深刻的大洗牌。AI 凭借卓越的算力，能够替代大部分依赖体力的常规流水线式试错实验。在日常科研实践中，化学家往往受限于自身专业的知识壁垒（例如不精通高分子物理），但通过与具有全学科视野的 AI 大模型进行持续“对话”，AI 能够瞬间调用跨学科的经典物理公式与宏观边界条件，带来密度极高的信息流与跨界启发，极大提升了从微观结构设计到宏观性能输出的研究效率。

第三章 生命系统中的物理法则：相变假说与意识涌现

如果说多相催化是无机物质形态的复杂衍生演化，那么生命科学则是宇宙中已知最为精密的物质科学。有学者直言，生命科学本质上就是“最复杂的物质科学”。细胞活动的底层逻辑毫无例外地建立在微观化学反应、分子间相互作用力、电荷转移以及热力学熵与能量的持续博弈之上。然而，生命系统内部的层级脱耦效应，远比无机物理世界更加显著且深奥。

3.1 从还原论困境到凝聚体（Condensates）理论

生命科学领域长期以来同样深受还原论思想的主导。例如，轰动一时的“人类基因组计划”曾让许多人乐观地认为，只要彻底完成全部基因测序，人类疾病乃至生命机理的所有问题都将迎刃而解。但事实证明，从微观分子序列到宏观系统功能之间存在着深不可测的鸿沟。现代神经科学对大脑组织的静态结构已经了如指掌：从神经细胞的形态、胶质支持细胞的分布，到底层突触连接以及神经递质的化学通讯机制，其还原论细节已经极度清晰。然而，当面临由百亿级神经元构成的复杂动态网络时，关于“学习、记忆如何确切储存”以及“意识究竟是如何从无机分子中涌现”的宏大命题，传统因果链条失效，常常陷入大量无法被当前技术验证的碎片化假说（Hypothesis）之中。

在此困境下，“相变”（Phase Transition）这一发源于物理学核心的理论概念，正在重塑生物学与脑科学的前沿范式。在细胞生物学中，相变常被称为“凝聚体”（Condensates）形成。最新的前沿交叉项目，如南科大深圳医学院与南京大学张晨宇教授团队探讨的“生命活动的涌现机制”，正致力于证明这样一个前卫的假说：大脑意识的涌现，极有可能正是始于微观突触（Synapse）在物质分子层面发生的动力学相变。当神经细胞进行高频信号通讯时，特定生物大分子的局部浓度突破临界值，从而发生物理层面上的液-液相分离（LLPS），形成具

有高度组织原理的集体效应。这种微观相变，在底层物理逻辑上与张广铭教授阐述的晶体对称性破缺原理如出一辙，构成了生命认知过程的物质基础。

上科大的罗振革教授，是该交叉领域的领军人物。其领衔的神经发育与再生研究团队，深入探究了自噬诱导在脊髓损伤后稳定微管及促进轴突再生的微观机制，在顶级期刊发表了大量标志性成果。罗教授的科研视角高度契合“生命即复杂物理大系统”的理念，呼吁建立跨越分子细胞神经科学与系统神经科学之间的宏大组织原则。

3.2 对称性破缺与生物智能的演化

在更宏大的生物演化尺度上，物理学的对称性破缺理论同样为“智能产生”提供了绝佳的哲学解释。生物学界关于智能萌芽存在一个极其生动的假说：远古低等生物在进化过程中“长出两只眼睛”的节点，正是真正智能开始涌现的元年。

在此之前，无眼或单眼生物对外界刺激的感知是模糊且高度对称的。而当双眼出现，其视野首次有了严格的左右之分，平移与镜像对称性宣告破缺。生物体开始面临环境生存中的两难选择（例如：左边感知到天敌威胁，右边感知到食物资源）。在无数次生死攸关的选择与对称性破缺的训练下，生物大脑和运动神经被迫进化出极其复杂的特征边缘信息整合机制与因果判定逻辑，从而在极其漫长的演化周期中，渐渐涌现出早期的生物智能。

面对如此高维度的生命演化现象，未来认知科学的终极突破，必须依赖大模型和 AI 在跨尺度特征提取上的强大能力，将底层物理分子信号通路与极宏观的意识形态建立起精确的数据映射。这也解释了上科大博士后何佳欢，虽然面临使用大量蛋白质预测与文献总结大模型带来的“信息过载”与“AI 幻觉焦虑”，但依然坚定地通过跨学科学习，试图在人类生殖医学（精卵结合演化）这种极度复杂的生命起源课题中，寻找人脑与 AI 协作的安全边界与科学突破口。

第四章 开放宏观系统中的社会经济模拟与数字孪生

相较于可以严格界定边界条件的物理系统与生物细胞系统，人类社会的经济大系统属于极其复杂的“开放系统”。它不仅伴随着物质与能量的流动，更充斥着海量的高频信息交互与异质性极强的人类主观心理博弈。

4.1 局部微观与全局宏观的合成谬误

经济学家在尝试理解社会系统时，早年亦大量借鉴物理学的方法，试图构建动态随机一般均衡模型（DSGE）等工具。然而，由于社会系统的变量极为庞杂，且经济学至今未能完全搞清系统最底层“脑细胞”（即微观个体的真实行为学特征与复杂协作机理），导致传统理论模型在面临真实政策推演时经常失效。

著名经济学家朱晓冬以中国基层的精准扶贫与乡村振兴政策为例，生动揭示了开放系统中的传导悖论与“合成谬误”（Fallacy of Composition）。在大量的实地调研中发现，尽管政府向农村注入了大量资金与科技资源，极大提升了局部的农业机械化水平与规模化生产效率，但许多农民最终依然未能实现收入的大幅增长。

导致这一现象的根本宏观机制并非农业技术落后，而是由于户籍制度与城乡壁垒未彻底打通，庞大的富余劳动力要素被硬性“锁定”在土地上，无法顺畅地向二三产业与城市流转分流。在这样一个要素缺乏自由流动的复杂系统中，如果单纯考察过去的数据，某个特定村庄率先利用科技提高生产率，在局部封闭市场内确实能获得超额利润；但如果推行全局性政策，导致所有村庄的生产率同时大幅提升，宏观系统便会发生结构性相变。海量的同质化高产农产品导致总体供给过剩，极致的宏观内卷推动市场价格坍塌。最终，微观个体的理性生产效率提升，加总后反而导致了所有参与者均不赚钱的宏观非理性结果。

4.2 基于 AI 大模型的数字孪生与全球治理前沿

上述案例表明，宏观政策所引发的复合系统效应远比简单的因果想象复杂。在未厘清跨层级耦合原理之前，推行强力干预政策往往会带来不可预知的全局性结果。

正因如此，传统依赖过去局部统计数据进行线性外推的治理模式将走向式微。学界当前高度寄望于利用 AI 大模型强大的复杂系统模拟与高维数据处理能力，构建中国乃至全球经济系统的“数字孪生”系统。通过在虚拟环境中设置具备异质性行为特征的智能体，政策制定者可以在实施真实政策前进行大规模计算机仿真与极端压力测试，提前预警系统演化方向及潜在的传导危险。在这一方向，上海科技大学教育、创新与可持续发展中心（CEISD）正积极布局，招募 AI 科学家、领域科学家和计算机科学与数据科学背景的工程师，推进创新探索。

第五章　AI 在工程、临床与科研设备管理中的应用涌现

理论界对于相变与衍生论的探讨，正在渗透并重塑现代工业工程、医疗系统与大型高校科研管理的实务操作逻辑。

5.1 从静态孤立到集群涌现的工业系统设计

在传统的创意工业设计领域，研发思路多局限于对单一“物”的静态微观打磨。但随着人工智能与物联网技术的成熟，系统所面临的物理当量正发生剧变。以自动驾驶出租车（Robotaxi）研究为例，当道路上仅仅运行少量测试车辆时，其涉及的交互场景极其有限且可被人工穷尽预设。然而，创艺学院杨继栋教授指出，一旦在现实交通网络中同时投入 10 万辆 Robotaxi，当这种庞大数量级达成时，整个交通系统就会自发涌现出在单车驾驶模拟中根本无法预见的海量、极其复杂的群体会车场景与极端路况碰撞。

在这种高维复杂度下，传统的基于旧系统框架打补丁的设计方法无法实现颠覆性创新（不破不立）。现代交互设计研究正在将“涌现（Emergence）”作为一种主动的方法论指导学生。通过在 AI 系统中主动寻找并注入新的高维扰动变量，研究者通过模拟推演，使系统在更高级别自发暴露出隐藏的新需求，进而提出前瞻性的群体解决方案 。

5.2 医疗系统的智能化重塑与数字引流

在医疗卫生这一关乎国民基础民生与生命科学落地的巨大开放系统中，AI 同样正在底层重塑整个行业的运营生态。具有丰富医院管理运营经验的专家黄玉琼观察到，AI 已经真切地改变了患者的就诊习惯与公立医院的精细化日常运营。

当前，包括微医、阿福等在内的数字化线上平台，已经能够利用智能化手段，自助完成高达 80% 的慢性病常规复诊与配药流程。同时，资本市场（如字节跳动巨资投入 60 亿）正全力构建以人工智能驱动的

“AI 医院”新业态。这种基础医疗需求向互联网端的大规模转移，使得传统的线下公立三甲医院面临严峻的“系统性挑战”与患者流量流失。为了破局，顶尖大学临床研究中心在肩负高精尖科研的同时，须运用先进的 AI 工具进行数字化转型。其核心战略目标是利用智能化手段，将庞大且宝贵的慢病与复诊患者群高效粘留在医院的闭环系统中，将连贯的生命体征长期随访数据，转化为驱动下一阶段医学大模型科研创新的核心数据源泉 。

5.3 科研共享平台的跨界耦合与“AI 荐仪”

在支撑科学研究的高校基础设施建设方面，大型仪器设备的共享管理同样呈现出脱耦与耦合的复杂生态。宏观层面上，管理者须具备顶层统筹学科发展的视野，以确保高精尖设备能够最大化地支撑教学科研体系；而在微观层面上，每一台精密单台设备又需实施严格的精细化维护标准。

在该领域拥有丰富经验的上科大设备资产处主管刘丹指出，大型仪器的共享过程绝非简单的硬件出借，而是一个如同生命体般具备持续自组织（Sustaining）能力的演化系统。随着跨学科研究的日益频繁，研究者在推进前沿交叉课题时，往往由于深厚的学科壁垒，并不知晓某些尖端物质科学设备的存在与其确切用法。为了打破这种层级间的脱耦，设备管理处引入了前沿的“AI 荐仪”系统。该系统利用算法挖掘跨学科数据关联，为跨界研究者推荐共享系统中现有的相关仪器。通过这种宏观规则的统筹与 AI 算法的微观纽带作用，设备共享平台在使用过程中不断孕育、涌现出原先未曾设想的新兴科研趋势与丰富的个性化使用需求，充分展现了管理设计在促进知识跨界耦合中的巨大潜力。

第六章 科学真理的逻辑底线与“人类在环”的哲学审视

尽管人工智能在处理复杂开放系统、挖掘高维隐藏关联方面展现出惊人的计算天赋，但在严谨的科学探索前沿，对其边界的界定须保持冷静与克制。

6.1 严密逻辑的不可妥协性与 AI 幻觉危机

科学体系的终极底线在于两大核心：第一是结论的可严格证明性，第二是推导过程的严密逻辑链条。凡是无法被严格论证、因果逻辑断裂的主张，均不属于科学真理的范畴。核聚变专家章林文教授警告，资本市场对所谓“AI 无所不能”的狂热造神，在严肃科学家眼中是极不负责任的。

当前的生成式大模型在本质上仍建立在基于过往历史数据的概率整合与文本概率分布之上，并未掌握真实的物理现实与世界模型。来自物质学院的年轻学子俊楠敏锐地指出，在自然语言逻辑模型中，AI 常常将“数学上的文本相等”与“物理现实中的条件对等”混为一谈。例如，大模型可能在文本层面上认为“达到 100 摄氏度”与“水开了”是完全等价的并列概念。然而在严密的物理因果链中，前者是后者发生热力学相变的先决条件，二者具有严格的时间先后顺序，不可简单地倒置划等号。

正由于大模型缺乏底层物理机理的硬性约束，其在推演中不可避免地会产生严重的“幻觉”（Hallucination）。倘若人类丧失了独立的辨别能力，将重大科学真理的裁判权拱手让给仅凭概率输出的黑盒系统，一旦其在涉及国家战略或生命科学的基础机理上犯下致命错误，将给人类带来不可估量的灾难。

6.2 创造力的本源机制与反常识直觉

人类最顶尖的创造力是如何诞生的？德国大哲学家叔本华曾提出过

关于创造力“微光”的深刻理论。他认为，惊天动地的科学与艺术突破，往往源自人类大脑中长年累月存储的无数记忆符号，在某个特殊的极端直觉瞬间发生的随机碰撞与自组织。这种前沿的开辟，往往极度依赖老一辈科学家向年轻学子指出的那个“打破常识的微光闪烁点”。

然而，如果科研群体过度依赖 AI 的加权平均路径，由于大模型的语料库本身是对人类已有既定历史常识的极度拟合，其输出天然带有极强的“规范化”与“框架化”特征。长此以往，那些真正具有革命性、颠覆现有理论框架的反常识“微光点”，在概率模型中极易被当作无效的“数据噪点”而遭到算法的洗劫。

6.3 确保安全的“人类在环”机制

为了对冲这种风险，设计与科研团队探索出了一种利用 AI 的“反向策略”：在人机协同中，研究者主动告知 AI 什么是已知的常规解、什么是坚决不能触碰的设计死角，通过设置严格的约束条件，迫使大模型跳出常规概率分布，在认知边缘地带输出超出想象的备选方案。随后，人类利用自身的主观能动性去审视这些边缘方案，从中汲取打破自身认知壁垒的逆向灵感 。

古希腊哲学家苏格拉底有名言：“我唯一知道的，就是我一无所知”。在浩瀚的宇宙规律与精妙的生命机理面前，人类现有的认知依然充满庞大的黑箱。因此，科学界的共识在于，目前阶段的 AI 不能被视为拥有独立灵魂与自我决策能力的物理“智能体”，而应被定义为高效的思考辅助伙伴。在整个科学发现的研究链路中，“人类在环”（Human-in-the-loop）是不可动摇的底线原则。人类专家须亲自站在回路的核心节点，掌控对齐与治理的主导权，并承担最终科学结论的判定责任。唯有如此，方能确保人工智能的发展路线图沿着“通过安全设计（Safe-by-Design）造福人类”的轨道前行。

结论

在从物理学的太初宇宙演化（太素）迈向万物运行数理法则（逻各斯）的伟大征途中，人类认识论经历了从经典还原论向现代衍生论的壮丽跨越。还原论为我们精准描摹了从夸克到大分子的微观基础积木，而衍生论则深刻揭示了在复杂层级交界面处，由于对称性破缺与相变现象所引发的新一轮“组织原理”重构。

进入二十一世纪，面对催化化学跨时空的理论断层、生命意识突触相变的深邃之谜，以及宏观经济政策极易诱发的系统性合成谬误，单纯依靠人类有限精力的手工试错和线性逻辑外推已触及天花板。以科学智能为代表的 AI 技术体系，凭借对高维非线性数据的超强特征提取与关联映射能力，正在改变自然科学与社会科学的研究范式。未来十至二十年，无疑是 AI for Science 在全球范围内引发学科重塑与知识重组的黄金发展期。

然而，无论技术如何狂飙突进，科学的内核——可证明性与因果逻辑底线不容妥协。面对大模型的内生幻觉与概率拟合特征，科研范式须坚守“人类在环”的审慎原则。人工智能是照亮未知黑箱的高聚光电筒，而决定探索方向、捕捉创造力颠覆“微光”并为最终真理担责的，唯有闪耀着人文理性与批判精神的人类心智。

联系我们：CEISD@Shanghaitech.edu.cn