

院士推荐语 “物理学家喜欢探索万物之理,养成万事都从第一性原理出发的思考习惯,也曾贡献了人工智能的早期探索。物理 for AI,是用物理学的已有的原理、概念、工具和思维方式,去理解、设计、优化人工智能系统;而 AI for 物理,则是用 AI 加速物理科研。罗老师在本文提出的物理第一性 AI,一方面包含了物理 for AI 的内涵;另一方面,对于复杂系统,特别是多体强关联复杂量子系统,对于远离平衡态的多体复杂系统,对于这样系统中的 emergent 现象等,物理学还远远不够完善,而这部分我以为与真正的 AI(不是基于黑匣子,靠 scaling law 出来的 AI)是密切相关的。因此,物理第一性 AI 以物理学基本理论为基础,将 AI 视为一种复杂物理系统,交叉融合其他众多学科,探索智能形成机制,也将物理学科推向新的高度。期待广大物理研究者与青年学子发挥专业所长,大胆突破,开拓创新,与此新兴未知领域共同成长。”

——朱邦芬院士(清华大学)

“各类基础物理原理与科学思想,始终是人工智能迭代演进的核心源头活水。本文提出物理第一性 AI 全新研究范式,兼具理论高度与前瞻视野。文中将量子理论视作拓展智能发展边界关键前沿方向之一,这条交叉融合之路的前沿探索方才起步,蕴藏着不容小觑的发展潜力,有待学界持续深耕、合力推进。文章打通物理学与智能科学的内在关联,从底层规律视角重新审视智能系统的构建逻辑,为跨学科前沿探索带来极具启发性的全新思考,值得大胆开拓,在扎实探索实践中开创一条立足物理底层逻辑、更具本原性的智能发展新路。”

——薛其坤院士(南方科技大学)

“当前以纯数据驱动为主的人工智能,面临理论体系碎片化、长程外推泛化不足等深层次难题。该文从对称性、统计力学、量子理论等物理视角出发,从底层原理阐释了物理范式能够助力可解释世界模型搭建和人工智能自主进化。论文表明基础物理理论可为迈向通用智能的前沿突破提供新思路。”

——段文晖院士(清华大学)

【编辑人语】罗迪副教授作为清华物理系 AI+Physics 与量子计算交叉领域的青年人才,在清华大学物理学百年系庆之际撰写此文并首次提出“物理第一性 AI”研究范式,前瞻性地探讨了基于物理规律和思想探寻超级智能的路径。相信此文将引领学界深入思考物理原理与人工智能的融合,为前沿研究提供富有原创性的学术洞见。

物理第一性 AI: 从物理通向智能

罗 迪^{1,2}

(¹ 清华大学物理系,北京 100084; ² 清华大学高等研究院,北京 100084)

摘 要 人工智能与物理学正进入深度交叉融合的新阶段。长期以来,物理学不仅为人工智能的发展提供了对称性、统计力学、量子理论等重要理论基础,也不断推动智能模型、训练优化与计算范式的创新;与此同时,随着人工智能系统复杂性的持续提升,其运行机理与演化规律也逐渐成为物理学关注和研究的新对象。在此背景下,本文提出**物理第一性 AI (Physics-first AI)**这一概念,将人工智能视为运行于客观物理世界中的复杂信息处理系统,从物理学的基本原理与研究思想出发,系统理解、设计并优化 AI 的模型结构、训练机制、计算方式与推理范式,形成一种新的人工智能研究范式。本文系统梳

收稿日期: 2026-07-08; 修回日期: 2026-07-18; 出版日期: 2026-08-22

通信作者: 罗迪, diluo@mail.tsinghua.edu.cn。

引文格式: 罗迪. 物理第一性 AI: 从物理通向智能[J]. 物理与工程, 2026, 46(5): 29-37.

Cite this article: LUO D. Physics-first AI: From physics to intelligence[J]. Physics and Engineering, 2026, 46(5): 29-37. (in Chinese)

理了人工智能与物理学交叉发展的理论脉络,并围绕结构先验、统计力学、量子理论和物理推理四个方面,阐述物理学推动新一代人工智能发展的理论基础与潜在路径,进一步讨论人工智能时代物理教育的发展方向。物理第一性 AI 有望为人工智能基础理论的发展提供新的物理视角,并推动科学智能、自主科学发现和未来通用人工智能的发展。

关键词 物理第一性 AI; 人工智能; 对称性; 统计力学; 量子理论; 物理推理; 物理智能模型; 量子优势

DOI 10.27024/j.wlygc.2026.07.08.is

PHYSICS-FIRST AI: FROM PHYSICS TO INTELLIGENCE

LUO Di^{1,2}

(¹ Department of Physics, Tsinghua University, Beijing 100084; ² Institute for Advanced Study, Tsinghua University, Beijing 100084)

Abstract Artificial intelligence (AI) and physics are entering a new stage of deep interdisciplinary integration. Physics has long provided foundational ideas for AI, including symmetry, statistical mechanics, and quantum theory, while continuing to inspire new model architectures, learning principles, and computational paradigms. At the same time, as AI systems become increasingly complex, their internal mechanisms, collective behavior, and emergent dynamics are becoming new objects of interest and investigation in physics. In this Perspective, we introduce **Physics-first AI**, a research paradigm that regards AI as a complex information-processing system embedded in the physical world and seeks to understand, design, and build intelligent systems from fundamental physics principles and modes of reasoning. Physics-First AI aims to provide a common foundation for model architectures, training mechanisms, computational processes, and reasoning paradigms through a unified physics perspective. We trace the historical interplay between physics and AI and discuss how structure priors, statistical mechanics, quantum theory, and physics reasoning may together shape the next generation of intelligent systems. We further examine how this emerging paradigm could reshape physics education and the training of interdisciplinary researchers in the AI era. Physics-First AI offers a route toward a physics-grounded theory and design framework for next-generation intelligence, opening new opportunities for scientific AI, autonomous scientific discovery, and ultimately artificial general intelligence.

Key words Physics-first AI; artificial intelligence; symmetry; statistical mechanics; quantum theory; physics reasoning; physics intelligence model; quantum advantage

2024 年诺贝尔物理学奖的颁布,为人工智能与物理学的跨界融合写下了里程碑式的一笔。该奖项授予约翰·霍普菲尔德(John Hopfield)与杰弗里·辛顿(Geoffrey Hinton),表彰他们“使利用人工神经网络进行机器学习成为可能的基础性发现和发明”,显示出物理思想与人工智能研究之间的深层联系受到学界高度认可。长期以来,大众普遍认为,人工智能主要聚焦于算法与工

程,物理学则致力于研究自然规律,二者相互分立。但随着技术迭代与理论深耕,这一认知正在被改变。

当代人工智能已经从传统的数据拟合模型逐步发展为能够完成复杂感知、生成推理任务的大规模学习系统,在训练动力学、参数耦合及集体行为等方面,呈现出典型的复杂系统特征,其运行规律、演化逻辑可以借助能量、信息、对称

性、动力学等物理概念与原理获得新的分析视角。与此同时,物理学不仅是人工智能的重要应用领域,而且持续为人工智能提供新的理论启发,涵盖模型结构、训练优化、计算范式、推理演化以及可解释性等多个方面。在此背景下,本文提出**物理第一性 AI (Physics-first AI)** 这一概念,将人工智能视为客观物理世界中的复杂信息处理系统,从物理学的基本原理与研究思想出发,系统理解、设计并优化 AI 的模型结构、训练机制、计算方式与推理范式,为下一代人工智能与科学智能的发展提供新的理论路径。

1 理论溯源: 物理与人工智能的百年交汇

人工智能与物理学的融合并非近年的新兴趋势,其理论渊源可追溯至百年来的科学理论积淀。统计力学、量子力学与复杂系统等基础物理理论,早已为现代人工智能的发展埋下理论伏笔,为智能计算提供了重要的概念基础与思想源泉。现代机器学习中的概率建模、能量模型与优化方法,其重要思想可追溯至玻尔兹曼、吉布斯等物理学家在 19 世纪后半叶奠定的统计力学体系。统计力学建立了微观状态与宏观规律之间的联系,以玻尔兹曼分布刻画系统微观状态的概率分布,并通过配分函数对所有可能状态进行归一化,从而描述系统的统计性质及其平衡行为。围绕概率分布、熵、配分函数和自由能等核心概念,统计力学形成了一套研究复杂系统的统一理论框架,也为后续神经网络模型的发展提供了重要理论基础^[1]。如今,机器学习中的能量模型、配分函数计算、对数似然估计以及变分推断等方法,都与统计力学框架具有密切的理论联系。从统计物理的视角看,人工智能的学习过程可以理解为系统在数据与任务约束下不断降低不确定性、优化能量-熵权衡的动态过程,自由能为分析这一过程提供了重要的理论视角。自由能最小化思想也被用于理解认知与学习过程^[2]。这种基于物理基本规律理解和设计学习系统的思想,体现了物理第一性原理在人工智能中的应用,也为机器学习模型的构建、训练与解释提供了具有物理意义的理论视角。

在经典神经网络的发展过程中,物理学提供了重要理论基础。Hopfield 网络、Boltzmann 机等经典神经网络模型直接源于伊辛模型和自旋玻

璃统计力学,而这些模型对现代神经网络的发展产生了重要影响^[1]。早期研究借鉴自旋系统的能量函数构建神经网络,将神经网络的记忆与检索过程描述为向能量极小值对应的稳定状态演化,建立了统计物理与神经网络信息处理机制之间的重要联系。后续研究进一步引入自由能、玻尔兹曼分布和变分推断等思想,推动了概率模型与能量模型的发展。不止于此,各类经典物理理论进一步拓展、完善了神经网络的设计体系:对称性理论推动等变神经网络的迭代演进^[3],为模型结构化设计提供严谨的约束依据;微分方程与物理动力学不断启发神经网络架构和生成模型的发展,例如,主流扩散模型借鉴非平衡热力学与随机扩散过程的思想设计其生成机制^[4];重整化群和相变理论也被用于理解深度神经网络的尺度行为、能力涌现^[5,6]等现象,为相关理论研究提供了新的分析视角。

与此同时,统计力学、复杂系统理论与量子力学的发展,深化了人们对信息与计算物理基础的认识,也为探索人工智能新的理论与计算范式提供了重要启发。薛定谔^[7]在《What is Life?》中探讨了如何从物理规律出发理解生命系统,推动了物理学与生命科学的交叉,并对后来关于复杂生命系统、信息与物理关系的研究产生了深远影响。兰道尔原理^[8]进一步揭示了信息处理的物理属性,指出信息擦除等逻辑操作具有不可忽略的热力学代价,表明信息处理必须依托具体的物理过程实现,因此计算能力与能耗最终受到其底层物理实现的约束。量子力学的发展则进一步拓展了人类对信息与计算的认识:冯·诺依曼建立了量子力学的系统数学框架,对量子态、测量与熵等概念进行了严格的数学表述,为后续量子信息科学的发展奠定了重要理论基础。费曼^[9]进一步指出,经典计算机在模拟一般量子系统时面临显著的计算困难,并提出利用量子系统模拟量子物理过程的思想,成为量子计算发展的重要理论起点之一。此后,量子计算进一步拓展至机器学习领域,推动了量子机器学习等交叉方向的发展,并为探索新的数据表示、采样与优化方法以及特定问题上的潜在计算优势提供了新的理论与技术路径。

2 人工智能的新发展: 物理范式的介入

过去十年是人工智能爆发式发展的黄金时

期,从 AlphaGo 突破围棋博弈壁垒^[10]、AlphaFold 破解蛋白质结构预测难题^[11],到图文生成模型实现高精度内容创作,大模型表现出一定的跨任务推理与自然语言交互能力,人工智能已经从专用模型逐渐发展为具备一定通用表示、生成和推理能力的基础模型。人工智能不再是辅助科研、生产的算法工具,而是演变为全新的知识组织方式与世界建模体系,深刻改变着人类认知世界、改造世界的方式。

人工智能高速演进之际,以数据驱动和工程迭代为主的发展范式也逐渐面临一系列深层挑战。其一,现有智能体系中的感知、生成、推理、决策与物理世界建模等能力仍缺乏统一、自洽的理论框架,如何借鉴物理学追求统一基本规律的研究思想,建立能够系统理解和描述智能底层运行机制的理论体系?其二,当前人工智能的发展在很大程度上依赖模型规模、数据规模与计算资源的持续扩展,但这一发展路径面临数据、算力、能耗和效率等多重约束,如何从物理第一性原理出发探索新的模型结构、优化目标与计算范式,为新一代人工智能开辟更加高效、可持续的发展路径?其三,当前人工智能仍高度依赖大规模数据与训练过程,在小样本泛化、因果推理、知识外推、原创发现、自进化与持续学习等方面仍存在重要挑战,如何借鉴物理学中从有限观测提炼普适规律、利用对称性与因果约束建立可外推模型的研究范式,推动 AI 从数据驱动的模式学习进一步走向规律发现、知识创造与自主迭代,进而探索迈向超级智能(artificial superintelligence, ASI)的可能路径?

物理学为人工智能的发展提供了具有潜力的系统性理论框架,可从结构、优化、计算、推理四大维度,搭建起物理第一性 AI 理论体系。在结构层面,物理学的对称性、守恒律、几何约束、因果结构和运动方程等规律定义了 AI 模型的基础架构先验。从平移、旋转等基础时空对称到置换、规范等更一般的对称,物理规律为模型连接方式、参数共享机制、变换不变性设计提供了核心准则,从而降低人工智能架构设计对经验性选择与反复试错的依赖。在优化层面,统计力学中的自由能、相对熵与变分原理,以及动力学与最优控制理论,为理解多类人工智能训练与优化问题提供了统一的理论视角。机器学习目标可以从能量、熵与信息之间的权衡视角加以理解,从而为人工智能的训练动

力学与优化机制提供具有物理意义的分析框架。在计算层面,量子计算与其他新型物理硬件,有望在特定任务中拓展传统计算边界。依托量子力学的物理特性,有机会突破经典算力瓶颈,拓展 AI 的应用场景与能力边界。在推理层面,物理思想有望打破传统人工智能对数据拟合与经验模式的过度依赖,提升 AI 在少样本推理、规律发现与创新探索方面的能力。同时,借助物理研究范式,AI 可形成“观测-假设-推演-验证-修正”的闭环学习过程,并探索智能系统持续学习、自我修正与自主进化的新路径。

结合结构、优化、计算与推理四个核心维度,本文提出物理第一性 AI(Physics-first AI)这一概念。物理第一性 AI 涵盖但不局限于已有的 Physics for AI 与 Physics of AI:前者侧重利用物理学改进 AI^[12,13],后者侧重利用物理学理解 AI^[14]。物理第一性 AI 将 AI 视为运行于客观物理世界中的复杂信息处理系统,而非单纯的数据拟合工具,以物理学的基本原理与研究思想为出发点,系统理解、设计并优化 AI 的模型结构、训练机制、计算方式与推理范式。物理第一性 AI 有望为通用人工智能的理论创新与技术突破提供新的物理视角,基于物理学底层逻辑构建新一代 AI。下文将系统阐释这一框架下物理学推动 AI 演进的内在机理与可能路径。

2.1 对称性与运动方程:构建具备结构化先验的 AI 模型

对称性是物理学中的重要基本原理^[15],也是机器学习架构设计中的重要原则。在物理学中,若系统的作用量在某连续群变换下保持不变,则该连续对称性对应一个守恒量。根据诺特定理,时间平移对称性对应能量守恒、空间平移对称性对应动量守恒、空间旋转对称性对应角动量守恒^[15],这一系列规律是自然系统稳定运行的核心底层逻辑。物理学的结构化研究思路,核心在于识别系统中的对称性与不变量,揭示复杂现象在不同变换下保持不变的内在规律。这一物理思想正好契合人工智能的架构优化需求。纯数据驱动的传统神经网络,往往忽略任务本身的物理对称性,需要大量数据与模型参数来学习其内在规律,泛化能力差、训练稳定性低。而依托物理对称性设计模型,本质是将自然规律前置嵌入网络结构,大幅降低模型的学习成本。其核心准则是实现神

神经网络的不变性与等变性设计:当输入数据发生对称变换时,模型输出保持不变或遵循对应变换规律。

当前主流 AI 架构的迭代升级体现了对称性思想的应用。卷积神经网络嵌入图像任务的平移对称性,通过参数共享实现平移等变性,大幅提升图像识别效率;图神经网络嵌入拓扑结构的置换对称性,实现节点置换等变性,适合社交网络、分子结构等数据;E(3)等变神经网络^[16]利用三维空间中的旋转与平移对称性,将分子和材料体系的空间变换规律嵌入模型架构,在分子建模、材料模拟等任务中展现出显著的性能优势。进一步,针对电子的反对称性^[17]和粒子物理的规范对称性^[18],学界已开发出遵循相应对称性的神经网络量子态,为人工智能模型求解量子多体物理体系提供新的路径。基于动力学与物理方程,学界还发展了物理信息神经网络^[19]、扩散模型^[4]、哈密顿量神经网络^[20]、拉格朗日量神经网络^[21]等前沿模型。哈密顿量神经网络通过神经网络学习系统哈密顿量,天然保留物理系统的辛结构与能量守恒特性,适用于轨道动力学、分子动力学等守恒系统建模;拉格朗日量神经网络直接学习系统拉格朗日作用量,适配带约束的广义坐标系统,可通过模型结构内生满足物理规律。与此同时,将物理运动方程直接融入神经网络计算,并依托光学、热力学、电子等多种物理硬件实现高效推理与训练,有望进一步推动神经形态计算(neuromorphic computing)的发展^[22]。

这一系列研究形成了物理赋能神经网络架构的设计思路:先识别任务对应的对称变换规律,确定模型所需的不变性、等变性特征,将物理结构前置嵌入模型,减少对人工正则化约束的依赖。这种设计思路让 AI 模型具备天然的物理一致性、更低的数据依赖性、更强的外推泛化能力。进一步,面向具身智能与多模态世界模型,仅依赖大规模数据拟合未必能够充分掌握真实世界的物理规律,智能系统还需要学习环境状态、动力学演化及行动后果之间的内在联系^[23]。物理学中的微分方程是描述世界演化的基本语言,其高精度求解是准确模拟与预测的重要基础^[24]。物理建模与仿真生成的动力学数据也可为世界模型训练提供重要支撑。事实上,物理学的核心正是研究世界如何演化以及支配这种演化的基本规律,这与世

界模型的目标具有天然一致性。由此,我们可以进一步探索“物理智能模型”(physics intelligence model),使其能够内化、学习与发现对称性、守恒律、运动方程、因果结构与有效理论等物理规律,并将其贯穿于感知、表示、仿真、预测与推理等机制中,发展以物理规律为内在结构的新型智能模型范式。

2.2 统计力学:提供 AI 训练优化的底层机制

统计力学的自由能变分原理,为监督学习的数据拟合、生成模型的分布重构、强化学习的奖励优化、大模型的对齐约束提供了统一的优化框架。自由能是统计力学的核心物理量,刻画了热力学系统中能量与熵之间的竞争关系,结合变分原理,可以将系统能量、熵、概率分布差异纳入共同分析框架。其核心物理内涵是:在适用的热力学条件下,可热化系统的自发演化趋向于降低相应自由能,并在充分弛豫后趋于自由能极小的热力学平衡态。

这一物理原理与机器学习的优化训练在数学形式上具有高度一致性。机器学习的最大似然训练,核心是最小化经验负对数似然,其数学本质等价于最小化数据真实分布与模型拟合分布的相对熵散度(Kullback-Leibler divergence, KL divergence),从而实现分布匹配。生成模型中的变分推断与证据下界(evidence lower bound, ELBO)优化,在数学结构上与统计物理中的变分自由能原理密切相关。以变分自编码器为例,其优化目标包含数据重建项与 KL 散度正则项,用于平衡数据拟合与潜变量分布约束;从变分推断的视角看,最大化 ELBO 等价于最小化相应的变分自由能^[25]。这一理论体系为监督学习、生成模型和强化学习等多类 AI 训练范式提供了统一的理论视角。在强化学习中,“控制即推断”(control as inference)^[26]将策略优化与概率推断联系起来,最大熵强化学习的最优策略在相应条件下呈现玻尔兹曼形式。KL 正则化策略优化进一步体现了期望奖励与策略偏离参考分布之间的权衡,其数学结构与统计力学中的自由能优化高度相似。

在大模型微调与对齐领域,基于人类反馈的强化学习(Reinforcement Learning from Human Feedback, RLHF)^[27]和直接偏好优化(Direct Preference Optimization, DPO)^[28]等方法,也可在 KL 正则化的偏好优化框架下获得统一理解。

以 RLHF 为例,其优化目标是在追求更高偏好奖励的同时,借助 KL 散度抑制策略相对预训练参考模型的过度漂移。从统计物理角度看,这类“奖励—KL 约束”优化与自由能最小化具有相似的数学结构:奖励可类比为负能量,KL 项刻画策略偏离参考分布的代价,其权重在形式上对应温度相关参数,用于调节奖励提升与参考分布约束之间的平衡,体现了统计力学中能量与熵竞争所形成的自由能优化思想。此外,重整化群、涨落耗散定理、平均场理论、相变、对称性结构与随机动力学等统计物理概念与理论,也为理解复杂智能系统中的多尺度表征、集体行为涌现、动力学转变与标度律,以及控制优化提供了重要启发^[5,6,14,29-31]。

2.3 量子理论:有望突破经典 AI 的表示与计算瓶颈

经典 AI 主要建立在经典信息表示与经典计算架构之上,随着模型规模和任务复杂度持续增长,逐渐面临上下文长度、计算与存储开销、长程信息有效利用以及组合泛化等方面的挑战。量子力学中的叠加、纠缠、干涉与相位等独特物理性质,为人工智能探索新的信息表示与计算范式提供了可能。量子计算的发展有望为未来 AI 提供新的硬件平台与算法模型。不同于经典比特的 0 或 1 状态,量子比特可以处于 0 和 1 的相干叠加态,并通过量子纠缠与干涉进行信息处理,从而有望在特定计算问题上实现超越已知经典算法的量子优势(quantum advantage)。

量子理论带来的潜在计算与表示优势体现在多个维度^[32]。其一为表示优势,量子态依托高维希尔伯特空间和量子力学特性,为表征高维数据中的非经典相关性和复杂特征结构提供了新的途径^[33];其二为线性代数加速^[34],量子相位估计、量子奇异值变换等量子算法,可显著降低特定矩阵运算的计算复杂度;其三为统计与搜索加速,幅度估计、Grover 搜索、量子游走等算法能够在特定条件下加速期望值估计、搜索与采样等计算过程^[35];其四为优化与模拟加速,量子退火、量子近似优化算法等为组合优化提供了新的求解路径,在金融投资、物流调度等领域具有潜在应用^[36,37];量子模拟则能够直接实现复杂量子系统的演化,在量子材料、量子化学和药物设计等领域具有潜在优势^[38]。与此同时,量子算法^[39]在整数分解等问题上的加速也对现代密码学产生了重要影响。

量子理论对人工智能的潜在赋能并非局限于计算加速,还可能为信息表示、学习机制与模型设计提供新的理论与计算范式^[40-41]。量子计算依托叠加、纠缠和干涉等量子特性进行信息表示与处理,为机器学习提供了独特的表示与计算机制;量子态表示、张量网络、量子多体物理与量子信息等理论与方法,也可驱动经典人工智能的架构设计、训练优化与长上下文标度律等研究^[42-44]。在量子机器学习领域,对于量子实验和量子模拟产生的量子数据,量子学习器可以直接处理量子态的信息,在部分任务中展现出可证明的样本或计算复杂度优势^[45];进一步利用量子计算处理经典数据、加速经典机器学习任务,也是当前量子计算与人工智能交叉研究的重要前沿方向。

2.4 物理推理:通向 AI 小样本学习创新与自进化

物理学的学习范式与科研思维,在小样本洞察、既有认知突破与动态迭代方面具有鲜明特点,为当前以大数据和规模扩展为主的 AI 发展范式提供了不同思路。物理学家善于从习以为常的普遍现象中洞察本质,从有限信息与少量观测中捕捉核心规律。认知的高度并非取决于知识与样本的数量堆叠,正如精读百本典籍的深度思考,往往胜于泛读千本书籍的浅层积累。依托第一性原理、对称性约束、等价转换、极限分析、量纲分析等数理推演,物理学能够从有限观测中凝练规律、构建模型、推演未知。这类有限信息下的科学推理与启发式问题求解,也在波利亚和杰恩斯的研究中得到探讨^[46,47]。纵观物理学发展历程,相对论、量子力学等重大理论革新,都体现了对既有经验与传统认知的突破;同时,物理理论通过“观测—假设—推演—验证—修正”不断迭代完善,为构建具备小样本学习、创新发现与持续进化能力的新型智能提供了重要的科学范式。

这种物理认知与科研范式,使 AI 能够从数据中主动发现解析规律与守恒结构^[48-51],有望促使 AI 超越以数据拟合为主的传统学习范式。当前主流 AI 高度依赖大规模数据,其能力往往受训练数据分布与既有知识边界制约,在小样本泛化、跨分布外推、反共识推理与原创性发现等方面仍面临挑战,模型能力的提升也在很大程度上依赖外部数据与反馈。引入物理第一性思维后,AI 可以突破对“数据越多、性能越强”单一路径的过度依

赖,探索“原理先验、小样本深挖、本质洞察”的学习模式,将因果关系、对称性等结构化知识融入学习与推理过程,提升小样本学习与跨分布外推能力。进一步借鉴物理学质疑既有假设、从有限现象中探寻深层规律的研究范式,AI有机会超越既有数据与知识框架,主动发现隐藏于复杂现象背后的非直观规律,促成突破既有认知的原创性科学发现。

近期评测研究表明,大模型在复杂物理推理、可靠计算与自主验证等研究能力上虽然取得进展,但是仍有较大提升空间^[52-56]。发展具备全流程自主研究、自我修正与多智能体协同能力的“AI物理学家”^[56],可能为更高水平自主科学智能乃至通用人工智能的探索提供新的研究视角。未来的“AI物理学家”有望将物理学的因果建模、理论推演与规律发现能力,与AI的大规模计算、搜索和智能体组织迭代能力相结合,使AI能够自主提出科学问题与假设,进行理论推导和数值模拟求解,开展实验验证,并根据反馈不断修正模型与理论,形成“观测现象-发问假设-建模推演-实验验证-反馈修正”的自主科研闭环。这样的AI不仅有望加速复杂科学问题的求解,还可提升推理的可验证性、可解释性与安全性,并实现自进化、持续学习、跨问题泛化与原创发现,为探索从通用人工智能走向更高水平的自主科学智能乃至超级人工智能(ASI)提供新的路径。

3 人工智能时代的物理教育:物理赋能智能的人才培养

AI的快速发展不仅正在重构技术体系,也对传统学科教育提出了深刻挑战。在大模型与AI智能体的推动下,传统编程的工作方式与能力要求正在被重新定义;随着自动定理证明与形式化验证技术的发展,数学研究与证明的工作范式也在发生变化。当计算机、数学等学科的研究与实践方式加速变革时,物理学教育同样亟需主动回应人工智能时代的新需求。

首先,物理教育需要实现工具能力的全面升级。新时代物理人才应主动掌握大模型、科学机器学习、自动微分、符号回归、物理仿真与智能科研工作流等前沿工具,同时夯实智能编程、并行计算、数值优化与数据处理等科学计算能力。教学

重点应从单纯“掌握编程工具”,进一步转向“运用计算与人工智能解决前沿物理问题”,推动AI深度融入凝聚态物理、量子信息、核聚变、宇宙学等研究领域,使其成为拓展物理科研能力的重要工具。

其次,物理教育需要培养兼具物理思维与AI理论创新和工程训练能力的复合型人才。物理学在对称性、变分原理、尺度分析和可解释建模等方面积累了独特而深厚的思想方法,为探索新一代AI提供了重要基础。AI时代物理教育应进一步打破物理、数学与计算机学科之间的壁垒,构建交叉融合的课程体系,将物理思想、数学基础与AI技术有机结合,培养能够理解、训练和优化AI模型并探索新型AI理论与架构的人才,鼓励学生从物理直觉和第一性原理出发,探索并突破现有AI范式的模型架构、优化方法、计算机制与推理方式。

最后,物理教育还应关注具身智能与世界模型的发展趋势。随着智能系统不断增强对空间、时间、因果关系、系统动力学以及环境反馈的建模能力,其发展与物理学研究真实世界规律的方法日益产生交汇。物理教育需要进一步强化实验训练,使得学生不仅具备理论推演与计算建模能力,还能够理解真实世界中的测量、实验设计与反馈控制,培养贯通“实验-计算-理论”的物理与人工智能交叉人才,为自主科学智能与物理智能等新兴方向提供人才基础。

物理教育当坚守以人为本,着力培育并提升人的物理核心素养;AI应融入科研探索,拓展人的思考与创造能力;凭借人积淀的物理认知突破智能发展瓶颈,构建人机协同、双向赋能的长效发展闭环。

4 结语

从百年理论溯源到当下的深度融合,物理学与AI的交叉已从早期的灵感借鉴,逐步走向系统化的方法论探索。物理第一性AI是人工智能突破固有瓶颈、走向通用智能的重要途径,也是物理学拓展研究边界、推动前沿科技发展的全新方向。未来,随着物理理论的持续深耕与人工智能技术的不断迭代,物理第一性AI有望重塑智能系统的设计、训练与发展范式,开启物理驱动的通用超级智能新时代。

参 考 文 献

- [1] HOPFIELD J J. Neural networks and physical systems with emergent collective computational abilities[J]. *Proceedings of the National Academy of Sciences*, 1982, 79(8): 2554-2558.
- [2] FRISTON K. The free-energy principle: a unified brain theory? [J]. *Nature Reviews Neuroscience*, 2010, 11(2): 127-138.
- [3] COHEN T, WELLING M. Group equivariant convolutional networks[C]//*International Conference on Machine Learning*. New York City: PMLR, 2016: 2990-2999.
- [4] SOHL-DICKSTEIN J, WEISS E, MAHESWARANATHAN N, et al. Deep unsupervised learning using nonequilibrium thermodynamics[C]//*International Conference on Machine Learning*. Lille: PMLR, 2015: 2256-2265.
- [5] MEHTA P, SCHWAB D J. An exact mapping between the variational renormalization group and deep learning[J/OL]. *arXiv*, 2014, *arXiv*: 1410.3831. <https://arxiv.org/abs/1410.3831>.
- [6] KAPLAN J, MCCANDLISH S, HENIGHAN T, et al. Scaling laws for neural language models[J/OL]. *arXiv*, 2020, *arXiv*: 2001.08361. <https://arxiv.org/abs/2001.08361>.
- [7] SCHRÖDINGER E. What is life? The physical aspect of the living cell[M]. Victoria, BC Canada: Rare Treasure Editions, 2025.
- [8] LANDAUER R. Irreversibility and heat generation in the computing process[J]. *IBM Journal of Research and Development*, 1961, 5(3): 183-191.
- [9] FEYNMAN R P. Simulating physics with computers[M]. Feynman and computation: exploring the limits of computers. USA: Perseus Books, 1999: 133-153.
- [10] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. *Nature*, 2016, 529(7587): 484-489.
- [11] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold[J]. *Nature*, 2021, 596(7873): 583-589.
- [12] CARLEO G, CIRAC I, CRANMER K, et al. Machine learning and the physical sciences[J]. *Reviews of Modern Physics*, 2019, 91(4): 045002.
- [13] JIAO L, SONG X, YOU C, et al. AI meets physics: a comprehensive survey[J]. *Artificial Intelligence Review*, 2024, 57(9): 256.
- [14] BAHRI Y, KADMON J, PENNINGTON J, et al. Statistical mechanics of deep learning. *Annual Review of Condensed Matter Physics*, 2020, 11(1): 501-528.
- [15] NOETHER E. Invariant variation problems[J]. *Transport Theory and Statistical Physics*, 1971, 1(3): 186-207.
- [16] BATZNER S, MUSAELIAN A, SUN L, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials[J]. *Nature Communications*, 2022, 13(1): 2453.
- [17] LUO D, CLARK B K. Backflow transformations via neural networks for quantum many-body wave functions[J]. *Physical Review Letters*, 2019, 122(22): 226401.
- [18] LUO D, CARLEO G, CLARK B K, et al. Gauge equivariant neural networks for quantum lattice gauge theories[J]. *Physical Review Letters*, 2021, 127(27): 276402.
- [19] KARNIADAKIS G E, KEVREKIDIS I G, LU L, et al. Physics-informed machine learning[J]. *Nature Reviews Physics*, 2021, 3(6): 422-440.
- [20] GREYDANUS S, DZAMBA M, YOSINSKI J. Hamiltonian neural networks[C]//*Advances in Neural Information Processing Systems*. 2019: 32.
- [21] CRANMER M, GREYDANUS S, HOYER S, et al. Lagrangian neural networks[J/OL]. *arXiv*, 2020, *arXiv*: 2003.04630. <https://arxiv.org/abs/2003.04630>.
- [22] MARKOVIĆ D, MIZRAHI A, QUERLIOZ D, et al. Physics for neuromorphic computing[J]. *Nature Reviews Physics*, 2020, 2(9): 499-510.
- [23] LECUN Y. A path towards autonomous machine intelligence version 0.9.2, 2022-06-27[EB/OL]. *Open Review*, 2022, 62(1): 1-62. <https://openreview.net/pdf?id=BZ5a1r-kVsf>
- [24] CHEN Z, MCCARRAN J, VIZCAINO E, et al. TENG: Time-evolving natural gradient for solving PDEs with deep neural nets toward machine precision[C]. *Proceedings of the 41st International Conference on Machine Learning*. Vienna, PMLR 235: 7143-7162, 2024.
- [25] BLEI D M, KUCUKELBIR A, MCAULIFFE J D. Variational inference: A review for statisticians[J]. *Journal of the American Statistical Association*, 2017, 112(518): 859-877.
- [26] LEVINE S. Reinforcement learning and control as probabilistic inference: Tutorial and review[J/OL]. *arXiv*, 2018, *arXiv*: 1805.00909. <https://arxiv.org/abs/1805.00909>.
- [27] OUYANG L, WU J, JIANG X, et al. Training language models to follow instructions with human feedback[C]//*Advances in Neural Information Processing Systems*. 2022, 35: 27730-27744.
- [28] RAFAILOV R, SHARMA A, MITCHELL E, et al. Direct preference optimization: Your language model is secretly a reward model[C]//*Advances in Neural Information Processing Systems*. 2023, 36: 53728-53741.
- [29] KWON C, AO P, THOULESS D J. Structure of stochastic dynamics near fixed points[J]. *Proceedings of the National Academy of Sciences*. 2005 Sep 13, 102(37): 13029-33.
- [30] ZIYIN L, CHUANG I, GALANTI T, et al. Formation of

- representations in neural networks [C]. In International Conference on Learning Representations. Vancouver, 2025 May 1 (Vol. 2025, pp. 68535-68563).
- [31] HABALA J, MARGOLIS G B, WANG T, et al. Agile asymmetric multi-legged locomotion; contact planning via geometric mechanics and spin model duality[J/OL]. arXiv preprint arXiv: 2602.09123. 2026 Feb 9.
- [32] HUANG H Y, CHOI S, MCCLEAN J R, et al. Vast World of Quantum Advantage[J]. Physical Review X, 2026, 16(3): 030501.
- [33] SCHULD M, KILLORAN N. Quantum machine learning in feature Hilbert spaces[J]. Physical Review Letters, 2019, 122(4): 040504.
- [34] GILYÉN A, SU Y, LOW G H, et al. Quantum singular value transformation and beyond; exponential improvements for quantum matrix arithmetics[C]//Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing. 2019: 193-204.
- [35] MONTANARO A. Quantum speedup of Monte Carlo methods[J]. Proceedings. Mathematical, Physical, and Engineering Sciences/The Royal Society, 2015, 471(2181): 20150301.
- [36] YARKONI S, RAPONI E, BÄCK T, et al. Quantum annealing for industry applications; Introduction and review [J]. Reports on Progress in Physics, 2022, 85(10): 104001.
- [37] BLEKOS K, BRAND D, CESCHINI A, et al. A review on quantum approximate optimization algorithm and its variants[J]. Physics Reports, 2024, 1068: 1-66.
- [38] GEORGESCU I M, ASHHAB S, NORI F. Quantum simulation[J]. Reviews of Modern Physics, 2014, 86(1): 153-185.
- [39] SHOR P W. Algorithms for quantum computation; discrete logarithms and factoring[C]//Proceedings of the 35th Annual Symposium on Foundations of Computer Science. Los Alamitos: IEEE, 1994, 124-134.
- [40] BIAMONTE J, WITTEK P, PANCOTTI N, et al. Quantum machine learning[J]. Nature, 2017, 549(7671): 195-202.
- [41] CERESO M, VERDON G, HUANG H Y, et al. Challenges and opportunities in quantum machine learning[J]. Nature Computational Science, 2022, 2(9): 567-576.
- [42] SToudenMIRE E, SCHWAB D. Supervised learning with tensor networks[C]//Advances in Neural Information Processing Systems 29. Red Hook: Curran Associates, 2016.
- [43] CHEN Z, DANGOVSKI R, LOH C, et al. Quanta; Efficient high-rank fine-tuning of llms with quantum-informed tensor adaptation [C]. Advances in Neural Information Processing Systems. Vancouver, 2024 Dec 16; 37: 92210-45.
- [44] CHEN Z, COMAS O, JIN Z, et al. L²M: Mutual Information Scaling Law for Long-Context Language Modeling [C]. Advances in Neural Information Processing Systems. San Diego, 2026 Apr 23; 38: 138466-506.
- [45] HUANG H Y, BROUGHTON M, COTLER J, et al. Quantum advantage in learning from experiments[J]. Science, 2022, 376(6598): 1182-1186.
- [46] POLYA G. How to solve it; A new aspect of mathematical method[M]. Princeton university press, 1945 Dec 31.
- [47] JEFFREYS H. The theory of probability[M]. OuP Oxford, 1998 Aug 6.
- [48] SCHMIDT M, LIPSON H. Distilling free-form natural laws from experimental data[J]. Science, 2009, 324(5923): 81-85.
- [49] WU T, TEGMARK M. Toward an artificial intelligence physicist for unsupervised learning[J]. Physical Review E, 2019, 100(3): 033311.
- [50] LIU Z, TEGMARK M. Machine learning conservation laws from trajectories[J]. Physical Review Letters, 2021, 126(18): 180604.
- [51] FANG Y L, JIAN D S, LI X, et al. AI-Newton: A concept-driven physical law discovery system without prior physical knowledge[J/OL]. arXiv, 2025, arXiv: 2504.01538. <https://arxiv.org/abs/2504.01538>.
- [52] QIU S, DENG J, DENG Y, et al. PRBench: End-to-end paper reproduction in physics research [J/OL]. arXiv, 2026, arXiv: 2603.27646. <https://arxiv.org/abs/2603.27646>.
- [53] WANG W, HUANG D, LI J, et al. Cmpphysbench: A benchmark for evaluating large language models in condensed matter physics [C]//International Conference on Learning Representations 2026. Rio de Janeiro: ICLR, 2026: 3388-3421.
- [54] MIAO T, JIN W, ZHANG M, et al. PRL-Bench: A Comprehensive Benchmark Evaluating LLMs' Capabilities in Frontier Physics Research[J/OL]. arXiv, 2026, arXiv: 2604.15411. <https://arxiv.org/abs/2604.15411>.
- [55] PAN H, ROGGEVEEN J, BERG E, et al. CMT-benchmark: A benchmark for condensed matter theory built by expert researchers[C]//International Conference on Learning Representations 2026. Rio de Janeiro: ICLR, 2026: 35522-35540.
- [56] DENG K, WANG X, DUAN G, et al. Towards verifiable and self-correcting ai physicists for quantum many-body simulations [J/OL]. arXiv, 2026, arXiv: 2604.00149. <https://arxiv.org/abs/2604.00149>.