

特色专题

天体物理人工智能: 基础模型与科研智能体

王瑜^{1,2}, 李亮³, 蔡荣根^{3*}

摘要 随着大视场巡天、高时间分辨率监测和多信使观测的快速发展,天文学正面临数据规模、数据类型与观测频率同步增长的新局面;与此同时,人工智能正从面向单一任务的专用模型,拓展至基础模型和以大语言模型为核心的科研智能体。综述了近5年上述方法在天体物理中的代表性进展,重点比较了专用模型、基础模型与科研智能体在训练目标、数据组织、迁移方式和科学角色等方面的功能边界。结果表明,专用模型在定义明确、数据分布稳定的任务中仍最为成熟有效;基础模型通过大规模异构数据预训练,已在图像、光谱和时间序列等跨任务迁移中展现出统一表征的潜力;而科研智能体尚处于早期探索阶段,其在文献归纳、工具调用和多步骤科研流程组织中的作用仍需审慎评估。通过进一步分析当前面临的数据质量、泛化能力、物理解释性、不确定性量化及可重复性等关键问题,指出未来应优先推进高质量数据基础设施建设、跨巡天外部验证规范、软件工具链固化以及交叉人才培养等方向。

关键词 天体物理;人工智能;基础模型;大语言模型;科研智能体;多模态学习

天文学研究正迈入一个全新的发展阶段。随着一系列大科学设施持续投入运行,大视场巡天、高时间分辨率监测、多波段观测,特别是多信使联合观测不断发展,天体物理研究越来越依赖对不同类型海量数据的统一分析与理解。与此同时,人工智能(artificial intelligence, AI)也在快速演进,其研究重点已由面向单一任务的模型训练,拓展至基于大规模数据的预训练和面向专业领域的模型适配^[1-4]。当代天文学的研究范式,正迫切寻求与人工智能技术发展趋势的深度融合。天体物理研究不仅积累了丰富的观

测数据,涵盖多类天体和多种物理现象,还具有明确但尚不完备的物理先验,同时形成了图像、光谱与时序数据并存的复杂数据生态。这些特点既界定了人工智能大模型在天文学中的重要应用场景,也为人工智能的发展提供了独特的科学环境。

理解当前人工智能的发展,首先需要区分几类人工智能模型在科研中不同的应用。以往广泛使用的小规模神经网络,通常是在相对固定的数据集上,围绕清晰定义的标签和目标函数进行训练,用于解决某一具体问题,例如光谱分类、红移估计或候选体识别^[5-6]。这类方法在很多成熟任务中依然最为有效,但它们更适合作为面向单一任务和特定数据的专用模型,而不是面向更广泛科学问题的通用基础。

基础模型(foundation model)通常在大规模、多来源、弱标签或无标签数据上进行预训练,以获得能够

1. 国际相对论天体物理中心(ICRA), 罗马 I-00185

2. 意大利国家天体物理研究所(INAF)阿布鲁佐天文台, 泰拉莫 I-64100

3. 宁波大学基础物理与量子科技研究院, 宁波 315211

收稿日期: 2026-03-25; 修回日期: 2026-06-28

基金项目: 国家自然科学基金卓越研究群体项目(A类)(12588101)

作者简介: 王瑜, 教授, 研究方向为高能天体物理、宇宙学和人工智能, 电子邮箱: yu.wang@icranet.org; 李亮(共同第一作者), 教授, 研究方向为高能天体物理、宇宙学和人工智能, 电子邮箱: liliang@nbn.edu.cn; 蔡荣根(通信作者), 研究员, 中国科学院院士, 研究方向为引力理论和宇宙学, 电子邮箱: cairg@itp.ac.cn

引用格式: 王瑜, 李亮, 蔡荣根. 天体物理人工智能: 基础模型与科研智能体[J]. 科技导报, 2026, 44(14): 37-44; doi:10.3981/j.issn.1000-7857.2026.03.00073

迁移至多种下游任务的数据表征^[3,7-8]。对于天文学来说,随着多波段、多时域和多模态观测数据持续累积,面对持续出现的新任务,研究者难以反复构建独立的数据表示、特征提取和训练流程。基础模型的潜力正在于,它有望把不同观测数据映射到更统一的表征空间,从而降低任务迁移成本,支持源分类、参数反演、异常发现以及多模态联合分析等一系列任务。

大语言模型(large language model, LLM)及由其驱动的科研智能体可参与文献归纳、知识整合、代码生成、工具调用、结果核验和科研记录等环节。当科研智能体与观测数据库、数值程序和分析软件连接后,其作用可由单步数据处理扩展至多步骤科研 workflows 的任务编排与协调^[9-11]。

需要指出的是,新兴方法并非对传统模型的简单替代。在许多具体任务中,尤其是在数据类型相对统一、特征明确的任务中,传统机器学习模型以及面向特定问题设计的神经网络,仍然具有很强的实用价值。因此,更符合当前天文学现实的图景,并不是“新模型取代旧模型”,而是不同层级方法在科研体系中的

分工逐步清晰:专用模型仍然是面向具体科学任务的落地主力,基础模型开始承担统一表征与跨任务迁移的功能,而大语言模型与科研智能体则进一步进入科研工作流层面。

通过综述近 5 年专用模型和基础模型的方法进展,讨论大语言模型与科研智能体在科研工作流中的作用,分析了相关发展趋势和主要问题。本研究中所称多模态主要指图像、光谱、时间序列、星表和文本等不同数据形态,多波段指不同电磁波段的联合观测,多信使则特指电磁辐射、引力波、中微子和宇宙线等不同物理信使的联合分析。

1 近 5 年天体物理人工智能的方法进展

为系统梳理专用模型、基础模型与大语言模型及科研智能体在科研体系中的不同定位,表 1 对比了这 3 类方法在天体物理应用中的主要差异,在实际研究中,3 类方法具有不同的优势和功能侧重点,共同构成天体物理人工智能的层级化工具生态。

表 1 专用模型、基础模型与大语言模型及科研智能体在天体物理应用中的主要差异

维度	专用模型	基础模型	大语言模型与科研智能体
训练目标	针对定义明确的单一任务进行优化	在大规模异构数据上开展自监督或弱监督预训练,学习可迁移表征	大语言模型学习语言和代码表征;科研智能体通过检索、工具调用和任务编排完成多步骤任务
数据组织	主要依赖单一巡天或特定任务的数据集	侧重跨巡天、跨模态和跨任务的数据整合与统一表示	连接文本、代码、文献、知识库、科学数据和外部工具
迁移方式	新任务通常需要重新训练或较大幅度微调	通过冻结编码器、线性探针或轻量微调迁移至下游任务	通过提示、检索、函数调用、记忆机制和轻量微调适配新任务
科学角色	完成定义明确的局部分析任务	提供可复用的数据表征和跨任务迁移基础	辅助组织多步骤科研流程并协调外部工具
主要局限	任务边界固定,跨数据集泛化能力有限	训练成本较高,评测和物理解释体系尚不成熟	存在幻觉、数值可靠性、版本漂移、可重复性和来源追溯问题

注:根据相关方法研究与代表性文献归纳^[3,5-11]。

1.1 专用模型:面向明确科学任务的高效方法

近 5 年,面向具体科学问题优化的专用模型仍是主要实用方法。例如,光学观测方向,类星体光谱分析是典型案例。Busca 等^[12]提出 QuasarNET,实现了基于深度神经网络的类星体光谱分类和红移估计。Rastegarnia 等^[13]通过卷积网络在斯隆巡天(Sloan Digital Sky Survey, SDSS)光谱上开展红移估计,强调

直接从光谱整体中学习局部与全局模式,而不是像传统方法那样主要围绕显著谱线进行模板式判别。随后, Moradi 等^[14]进一步将这一思路扩展到类星体、星系、恒星等的统一光谱分类。Liu 等^[15]以多任务卷积神经网络自动搜索并表征 Ca II 吸收线,把传统上高度依赖人工核验的流程推进至可扩展的自动化阶段;该方法在 SDSS DR7 和 DR12 数据中重新识别出 321 个

已知 Ca II 吸收体, 并从另外 381 条光谱中筛选出候选吸收系统。在高能天体物理中, Sahakyan 等^[16]利用梯度提升树对费米卫星数据库中的不确定型耀变体进行分类, 结果表明在小样本且带有缺失值的目录数据中, 仍然可以达到高于传统方法的准确率。随后, Begue 及其合作者把卷积神经网络引入耀变体宽波段辐射模型拟合, 先后构建了面向同步辐射-自康普顿、外康普顿以及质子同步辐射/混合模型的参数拟合网络^[17-19], 显著降低了传统自洽辐射模型的计算时间, 使快速宽波段拟合成为可能^[17-19]。基于上述工作建立的马卡良多波段数据中心(Markarian Multiwavelength Data Center, MMDC)把多时段多波段的数据获取和机器学习驱动的物理建模结合到同一平台, 发展了“数据访问—模型拟合—物理解读”的一体化工具链^[20]。专用模型在时域和高能瞬变领域能够嵌入观测流程, 快速服务新任务。Zhang 等^[21]提出基于迁移学习和多尺度卷积模块的伽马暴识别方法, 在标签有限的极目卫星(Gravitational wave high-energy Electromagnetic Counterpart All-sky Monitor, GECAM)数据上获得高精度结果, 并发现若干此前未识别事件; 其后续工作又通过数据增强与特征增强进一步优化了识别框架^[22]。在更高能区间, Li 等^[23]将机器学习用于高海拔宇宙线观测站(Large High Altitude Air Shower Observatory, LHAASO)的甚高能伽马观测中的背景抑制, 实现伽马射线事例与宇宙线强子背景的分类。

更前沿的趋势则是让人工智能参与到经验规律提炼和大样本物理量推断中。例如, Liu 等^[24]结合深度学习、量纲分析和符号回归, 从快速射电暴数据中提炼出一个分类方程; 其中所采用的符号回归思路与从深度模型中提取解析表达式的方法相联系^[25]。Lu 等^[26]利用以回响映射质量为标签的深度网络, 对 263604 个 SDSS-DR16 类星体给出黑洞质量估计; 在按天体划分的独立测试集上, 模型均方根误差约为 0.397 dex, 与回响映射质量约 0.35~0.5 dex 的内禀离散度相当。基于仿真的推断已成为复杂天体物理参数后验估计的重要方法^[27]。Dax 等^[28]建立了双中子星并合的机器学习推断框架, 可在约 1 s 内完成全部 17 个参数的后验推断; 与近似低延迟方法相比, 其 90% 可信天空区域面积的中位数缩小约 30%。

以上工作说明, 专用模型并未过时。相反, 它们

在近 5 年内扩展到了更复杂的数据形态和更高价值的科学问题之中。但从方法论上说, 这类模型大多仍然具有明确的应用边界。例如红移模型不一定能做分类, 伽马射线暴的模型不能直接应用于类星体的数据。对于未来的科研流程, 专用模型高效和精确的特征适合作为人工智能体系的“执行部件”。

1.2 基础模型: 统一表征与跨模态迁移

相比之下, 基础模型的目的不是把某一项任务做到极致, 而是构建可跨任务复用的统一表征。这一方向的前提首先是数据组织方式的改变。The Multimodal Universe 数据集把图像、光谱和时序天文数据以统一格式加以整理, 形成多达数亿观测、约 100 TB 规模的机器学习友好数据集, 并提供了代表性基准任务^[29]。这类工作的重要性在于为天体物理基础模型建立了公共数据基础, 服务于之后的“预训练”和“跨模态统一表征”。

在具体模型层面, 近年较有代表性的时域基础模型是利用约 50 万个合成超新星样本, 通过对比学习把光变曲线与光谱对齐在共同嵌入空间, 再基于 4702 个真实观测超新星样本进行微调^[30]。其方法学意义在于, 它展示了基础模型如何应用于“高质量光谱或标注样本稀缺, 而低质量观测或模拟样本相对丰富”这种典型的情景。对天体物理学来说, 这比单纯追求更高精度更重要, 因为未来多数大科学设施都将面临这种标注和模态不均衡的问题。

天文全模态网络 AION-1(Astronomical Omni-modal Network)代表了更接近通用天文基础模型的一种探索^[31]。其两阶段架构采用模态专用词元化, 并借鉴掩码自编码思想开展 Transformer 建模^[31-32]。模型在 Legacy Survey、超主焦相机(Hyper Suprime-Cam, HSC)、SDSS、暗能量光谱仪(Dark Energy Spectroscopic Instrument, DESI)和盖亚卫星(Gaia)等 5 个大型巡天上进行预训练, 覆盖超过 2 亿条恒星、星系和类星体观测数据^[31]。其意义主要体现在 2 个方面: 首先, 它把不同类型的天文观测数据纳入同一个模型进行统一处理, 这意味着研究流程有望从过去针对不同数据分别构建处理流程和分析模型, 转向优先学习统一的数据表征。其次, 它表明只需冻结预训练编码器, 并结合较为简单的下游适配模块, 就可以迁移到多类天体物理任务。这表明, 基础模型在天体物理研

究中正初步呈现出类似通用基础设施的作用。

中国团队提出的面向光谱数据的对比学习预训练框架 SpecCLIP(Spectroscopic Contrastive Language-Image Pre-training),围绕跨巡天、跨分辨率光谱对齐构建面向光谱数据的“窄基础模型”^[33]。其训练范式借鉴对比语言—图像预训练(Contrastive Language-Image Pre-training, CLIP)和简单视觉表征对比学习框架(Simple Framework for Contrastive Learning of Visual Representations, SimCLR)所代表的对比学习方法^[34-35],将郭守敬望远镜(Large Sky Area Multi-Object Fiber Spectroscopic Telescope, LAMOST)与 Gaia 的低分辨率光谱对齐至共同表征空间,并利用附加解码器保留不同仪器的特征信息^[33]。其重要性在于把“基础模型”落实到不同来源光谱统一表征这一可直接检验的问题上。

从数据模态看, AstroCLIP 通过对比学习将星系图像和光谱对齐至共享表征空间,可用于跨模态检索、红移估计、物理参数估计和形态分类^[7]; FALCO 则在 Kepler 光变曲线上开展自监督预训练,并迁移至恒星变源分类、表面重力估计和耀斑识别等时域任务^[8]。

综上,基础模型与专用模型的差异,并不在于其参数的多寡,而是能否在更大范围内减少重复建模与重复训练。对天体物理而言,基础模型最有价值之处在于为多源、多模态观测数据联合分析的科学任务提供一个共享基础。

1.3 大语言模型与科研智能体:科研流程的组织与自动化

面向天文学科研文本的领域模型方面, AstroLLaMA 利用 30 余万篇天文学论文摘要对 LLaMA-2 进行领域适配,代表了天文学专用大语言模型的早期探索。不过,领域适配并不等同于科学推理能力,其事实准确性、引文可靠性和工具调用能力仍需通过独立基准进行验证^[36]。

基础模型侧重学习可迁移的数据表征,大语言模型与科研智能体则用于组织文献、数据、代码和工具调用等科研环节。

关于大语言模型在天文研究中的早期使用, Fouesneau 等^[37]调研了 13 位天文学研究者在多个月内使用 ChatGPT、Copilot 等工具的实际经验,发现其

主要用途集中在论文写作、代码辅助和日常沟通等方面。该研究表明,现阶段大语言模型在天文学中的主要作用是辅助写作、编程和信息整理,提高日常科研效率,而非替代研究者的科学判断。

在更直接的科研应用中, Li 等^[38]将大语言模型用于样本稀少且噪声复杂的引力波识别任务。在仅使用 90 个激光干涉引力波天文台(Laser Interferometer Gravitational-Wave Observatory, LIGO)事件进行微调的设置下,模型的信号识别准确率达到 97.4%;进一步增加模拟样本并未改善其性能。这一结果提示,大语言模型在小样本和非高斯、非平稳噪声条件下具有探索价值,但其泛化能力仍需通过更多真实事件、不同观测运行期数据和独立模型加以验证。Stoppa 等^[39]展示了大语言模型在瞬变源分类中的应用潜力。仅通过 15 个示例和文字指令,模型就在 Pan-STARRS、MeerLICHT 和 ATLAS 等巡天数据上取得平均 93% 的分类准确率,并能够为每个候选体生成自然语言说明;但这种说明是否忠实反映模型的内部决策过程,仍需通过独立方法验证。

更进一步的探索来自 Wang 等围绕“AI 能否理解宇宙”开展的一系列研究。Wang 等^[40]通过使用多类天体物理数据对生成式预训练 Transformer(generative pre-trained transformer, GPT)模型进行领域微调,结果显示单一模型能够同时处理天体分类、伽马暴分类、类星体红移估计和黑洞参数推断等任务。在后续研究中, Wang 等^[41]进一步从直觉、因果关系和多模态处理等角度探讨了人工智能“科学理解”的含义。这一系列工作最有价值的地方不在于它们证明人工智能具备了某种类型的“科学理解”,更在于它们把问题从“能否再做一个更准的分类器”提升到了“能否构造统一的科学认知框架”层面。

智能体方法则进一步把大语言模型从“分析数据”推进到“组织研究流程”。多波段星系观测自动分析的自我改进型大语言模型智能体(Mephisto),把大语言模型、多智能体协作、树搜索,以及星系建模工具结合起来,让系统能够像研究者一样,围绕观测数据不断提出解释、调用外部程序进行检验,并根据结果反复修改判断,逐步收敛到较合理的物理图景^[42]。虽然这项工作目前尚处于探索阶段,但已清楚指明方向,相关方法正在由单步预测工具向能够调用外部工

具、整合先验并管理多步骤任务的科研智能体发展。

天体物理中的人工智能研究正经由大语言模型经历2方面变化。一方面,研究目标从早期侧重分类、检索和发现,逐步转向更高层次的科学分析,在复杂数据中提炼规律并参与科学推断。另一方面,人工智能介入的层次也正从训练模型处理科学数据,逐步扩展到科研流程本身,包括资料整理、假设生成、逻辑推理、工具调用、结果检验和总结结论。

2 发展趋势与主要问题

2.1 从模型架构转向数据与表征基础设施

从专用模型向基础模型的演进历程可以发现,真正重要的变化并不只是模型越来越大,更在于人们逐渐认识到,科研中大量重复投入的时间和精力,往往并不是投入在模型本身,而是在数据整理、不同类型数据的衔接、训练目标的设定,以及如何把数据变成模型能够有效利用的形式这些更基础的工作环节^[4-6,29,43-44]。因此,未来天体物理人工智能的研究能力,将更多取决于高质量公共数据集、统一的数据规范和可复用的通用表征。正如Wang等所强调,未来大科学装置的设计不能只考虑观测能力本身,应从设计阶段就把人工智能对数据质量、数据格式、可获取性以及后续分析流程的需求纳入整体考虑^[40]。

2.2 从预测性能转向物理解释与不确定性评估

对于天文学而言,当前的关键已不仅在于模型给出输出结果,而在于其能否帮助研究者理解结果为何成立、哪些证据支持这一判断、背后对应什么样的物理过程,以及这一判断在整个研究流程中是如何逐步形成的。正因如此,未来有价值的科研代理系统,不应只是更强的预测工具,而应当能够把数据分析、推理说明和工具调用组织到同一 workflow 中,清楚展示所依据的数据、假设和推理步骤,并在必要时完成外部检验。只有这样,人工智能才可能从提升效率的辅助手段,进一步发展为支持科学理解与科学发现的研究伙伴^[45-46]。

除物理解释外,不确定性评估应与点预测同时报告。对于分类任务,应给出概率校准误差、可靠性图和选择性风险;对于参数反演,应报告后验分布或置信区间,并区分由观测噪声引起的数据不确定性与由

模型结构、训练样本有限性引起的认知不确定性。面向跨巡天应用时,还应通过域外测试和校准漂移监测,评估数据分布变化对模型可信度的影响^[47]。

2.3 从单步工具转向可追踪的科研智能体

在科研场景中,大语言模型的价值不应只是局限于生成文本。更值得期待的方向,是其将文献阅读、数据处理、代码调用、数值模拟和不确定性分析组织到同一流程中,成为一个可以检查、可以追踪、也可以纠错的研究协调层^[11,42,48-49]。这意味着未来评价科学大语言模型的关键指标,也不应只是通用问答得分,还应包括以下方面:是否能稳定调用外部工具、是否能遵守物理约束、是否能给出可验证来源、是否能在版本更新后保持可重复性。换言之,天体物理真正需要的不是“会对话的通用模型”,而是可审计、可复现的科研工作流协调系统。

已有研究记录了大语言模型生成虚构或信息错配引文的问题^[50],并提出利用语义熵识别模型幻觉的方法^[51]。近期研究还指出,仅以准确率作为优化和评价目标,可能促使模型在不确定时进行猜测,从而产生看似合理但实际错误的答案^[52]。基于这些风险,建议科研智能体同时保留原始数据、文献标识符、代码和模型版本、工具调用日志及不确定性信息,并开展数字对象标识符(digital object identifier, DOI)或天体物理数据系统(Astrophysics Data System, ADS)校验、单位和边界检查以及独立程序复算。

2.4 从单项应用转向面向人工智能的科研基础设施

基础模型和大语言模型的建设难以由单个课题组独立完成,需要天文学、统计学、机器学习和软件工程等多个学科方向的长期协同合作。对未来的天文重大项目,人工智能相关能力不应只被看作附属支持,而应成为研究体系的一部分。这意味着,在项目组织上,需要把数据工程师和人工智能研究者纳入核心团队;在人才培养上,则需要建立更稳定的交叉培养机制,使研究者既能理解天体物理问题,也具备处理大规模数据和复杂模型系统的能力。

3 结论

近5年来,人工智能在天体物理领域的应用,已从面向单一任务的专用模型,逐步扩展至基础模型和

科研智能体。这3类方法并非简单替代关系:专用模型适合完成定义明确且数据分布相对稳定的任务;基础模型通过自监督或弱监督预训练提高表征复用和跨任务迁移能力;大语言模型与科研智能体则尝试连接文献、数据、代码和外部工具,参与多步骤科研流程的组织。现有证据表明,专用模型的应用最为成熟,天体物理基础模型已在图像、光谱和时间序列任务中显示出迁移潜力,而科研智能体在天文学中的应用仍处于早期阶段,尚不能替代研究者的科学判断。

后续发展应重点推进高质量数据与元数据建设、跨巡天外部验证、物理约束和不确定性评估、模型与工具版本固化,以及可追踪的端到端科研记录。只有当数据来源、模型输出、工具调用和人工干预均可检查、可复现时,人工智能带来的效率提升才可能进一步转化为可靠的天体物理研究成果。

参考文献 (References)

- [1] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436–444.
- [2] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems 30. Long Beach, USA: Curran Associates, 2017: 5998–6008.
- [3] Bommasani R, Hudson D A, Adeli E, et al. On the opportunities and risks of foundation models[J/OL]. *arXiv*, 2021: 2108.07258. [2026–01–21]. <https://arxiv.org/abs/2108.07258>.
- [4] Wang H C, Fu T F, Du Y Q, et al. Scientific discovery in the age of artificial intelligence[J]. *Nature*, 2023, 620(7972): 47–60.
- [5] Fluke C J, Jacobs C. Surveying the reach and maturity of machine learning and artificial intelligence in astronomy[J]. *WIREs Data Mining and Knowledge Discovery*, 2020, 10(2): e1349.
- [6] Huertas-Company M, Lanusse F. The dawes review 10: The impact of deep learning for the analysis of galaxy surveys[J]. *Publications of the Astronomical Society of Australia*, 2023, 40: e001.
- [7] Parker L, Lanusse F, Golkar S, et al. AstroCLIP: A cross-modal foundation model for galaxies[J]. *Monthly Notices of the Royal Astronomical Society*, 2024, 531(4): 4990–5011.
- [8] Zuo X X, Tao Y H, Huang Y, et al. FALCO: Foundation model of astronomical light curves for time domain astronomy. implementation and applications on Kepler data[J]. *The Astronomical Journal*, 2026, 171(1): 10.
- [9] Zhang Y B, Khan S A, Mahmud A, et al. Exploring the role of large language models in the scientific method: From hypothesis to discovery[J]. *npj Artificial Intelligence*, 2025, 1: 14.
- [10] Boiko D A, MacKnight R, Kline B, et al. Autonomous chemical research with large language models[J]. *Nature*, 2023, 624(7992): 570–578.
- [11] Schick T, Dwivedi-Yu J, Dessi R, et al. Toolformer: Language models can teach themselves to use tools[C]//Advances in Neural Information Processing Systems 36. New Orleans, USA: Curran Associates, Inc, 2023: 68539–68551.
- [12] Busca N, Balland C. QuasarNET: Human-level spectral classification and redshifting with Deep Neural Networks [J/OL]. *arXiv*, 2018: 1808.09955. [2026–01–21]. <https://arxiv.org/abs/1808.09955>.
- [13] Rastegarnia F, Mirtorabi M T, Moradi R, et al. Deep learning in searching the spectroscopic redshift of quasars[J]. *Monthly Notices of the Royal Astronomical Society*, 2022, 511(3): 4490–4499.
- [14] Moradi R, Rastegarnia F, Wang Y, et al. FNet II: Spectral classification of quasars, galaxies, stars, and broad absorption line (BAL) quasars[J]. *Monthly Notices of the Royal Astronomical Society*, 2024, 533(2): 1976–1985.
- [15] Liu Y, Li J, Gao L Q, et al. Deep learning for identification and characterization of Ca ii absorption lines: A multitask convolutional neural network approach[J]. *The Astrophysical Journal Supplement Series*, 2025, 276(2): 37.
- [16] Sahakyan N, Vardanyan V, Khachatryan M. Gradient boosting decision trees classification of blazars of uncertain type in the fourth Fermi-LAT catalogue[J]. *Monthly Notices of the Royal Astronomical Society*, 2022, 519(2): 3000–3010.
- [17] Bégué D, Sahakyan N, Dereli-Bégué H, et al. Modeling blazar broadband emission with a convolutional neural network. I. Synchrotron self-Compton model[J]. *The Astrophysical Journal*, 2024, 963(1): 71.
- [18] Sahakyan N, Bégué D, Casotto A, et al. Modeling blazar broadband emission with convolutional neural networks. II. external Compton model[J]. *The Astrophysical Journal*, 2024, 971(1): 70.
- [19] Sahakyan N, Bégué D, Casotto A, et al. Modeling blazar broadband emission with convolutional neural networks. III. Proton synchrotron and hybrid models[J]. *The Astrophysical Journal*, 2025, 990(2): 222.
- [20] Sahakyan N, Vardanyan V, Giommi P, et al. Markarian multiwavelength data center (MMDC): A tool for retrieving and modeling multitemporal, multiwavelength, and multi-messenger data from blazar observations[J]. *The Astronomi-*

- cal Journal, 2024, 168(6): 289.
- [21] Zhang P, Li B, Gui R Z, et al. Advancing gamma-ray burst identification through transfer learning with convolutional neural networks[J/OL]. arXiv, 2024: 2408.13598. [2026-01-21]. <https://arxiv.org/abs/2408.13598>.
- [22] Zhang P, Li B, Gui R Z, et al. An advanced identification method for gamma-ray bursts with data and feature enhancements[J]. *The Astrophysical Journal Supplement Series*, 2026, 282(2): 59.
- [23] Li J, Lü H K, Liu Y, et al. Application of machine learning to background rejection in very-high-energy gamma-ray observation[J]. *The Astrophysical Journal Supplement Series*, 2025, 276(1): 24.
- [24] Liu Y, Lu Y H, Moradi R, et al. Machine phenomenology: A simple equation classifying fast radio bursts[J/OL]. arXiv, 2025: 2512.04204. [2026-01-21]. <https://arxiv.org/abs/2512.04204>.
- [25] Cranmer M, Sanchez-Gonzalez A, Battaglia P, et al. Discovering symbolic models from deep learning with inductive biases[J/OL]. arXiv, 2020: 2006.11287. [2026-01-21]. <https://arxiv.org/abs/2006.11287>.
- [26] Lu Y H, SiTu H, Li J, et al. 263604 Supermassive black hole masses: Deep learning with reverberation-mapping calibration[J]. *Journal of High Energy Astrophysics*, 2026, 53: 100624.
- [27] Cranmer K, Brehmer J, Louppe G. The frontier of simulation-based inference[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(48): 30055-30062.
- [28] Dax M, Green S R, Gair J, et al. Real-time inference for binary neutron star mergers using machine learning[J]. *Nature*, 2025, 639(8053): 49-53.
- [29] Angeloudi E, Audenaert J, Bowles M, et al. The multimodal universe: Enabling large-scale machine learning with 100 TB of astronomical scientific data[C]//*Proceedings of Advances in Neural Information Processing Systems 37*. Vancouver, Canada: Curran Associates, Inc, 2024: 57841-57913.
- [30] Zhang G, Helfer T, Gagliano A T, et al. Maven: A multimodal foundation model for supernova science[J]. *Machine Learning: Science and Technology*, 2024, 5(4): 045069.
- [31] Parker L, Lanusse F, Shen J, et al. AION-1: Omnimodal foundation model for astronomical sciences[J/OL]. arXiv, 2025: 2510.17960. [2026-01-21]. <https://arxiv.org/abs/2510.17960>.
- [32] He K M, Chen X L, Xie S N, et al. Masked autoencoders are scalable vision learners[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE, 2022: 15979-15988.
- [33] Zhao X S, Huang Y, Xue G R, et al. SpecCLIP: Aligning and translating spectroscopic measurements for stars[J]. *The Astrophysical Journal*, 2026, 998(2): 189.
- [34] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//*Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*. Virtual Conference: PMLR, 2021: 8748-8763.
- [35] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//*Proceedings of the 37th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 119)*. Virtual Conference: PMLR, 2020: 1597-1607.
- [36] Nguyen T D, Ting Y S, Ciuca I, et al. AstroLLaMA: Towards specialized foundation models in astronomy[C]//*Proceedings of the Second Workshop on Information Extraction from Scientific Publications*. Kerrville: Association for Computational Linguistics, 2023: 49-55.
- [37] Fouesneau M, Momcheva I G, Chadayammuri U, et al. What is the role of large language models in the evolution of astronomy research? [J/OL]. arXiv, 2024: 2409.20252. [2026-01-21]. <https://arxiv.org/abs/2409.20252>.
- [38] Li Y X, Lu Y H, Liu Y, et al. Large language models for limited noisy data: A gravitational wave identification study[J/OL]. arXiv, 2025: 2512.04031. [2026-01-21]. <https://arxiv.org/abs/2512.04031>.
- [39] Stoppa F, Bulmus T, Bloemen S, et al. Textual interpretation of transient image classifications from large language models[J]. *Nature Astronomy*, 2025, 9(12): 1869-1878.
- [40] Wang Y, Zhang S R, Momtaz A, et al. Can AI understand our universe? test of fine-tuning GPT by astrophysical data[J/OL]. arXiv, 2024: 2404.10019. [2026-01-21]. <https://arxiv.org/abs/2404.10019>.
- [41] Wang Y. Reflections on “can AI understand our universe?” [J]. *International Journal of Modern Physics D*, 2026, 35(6): 2540010.
- [42] Sun Z C, Ting Y S, Liang Y B, et al. Mephisto: Self-improving large language model-based agents for automated interpretation of multi-band galaxy observations[J/OL]. arXiv, 2025: 2510.08354. [2026-01-21]. <https://arxiv.org/abs/2510.08354>.
- [43] Wilkinson M D, Dumontier M, Aalbersberg I J, et al. The FAIR Guiding Principles for scientific data management and

- stewardship[J]. *Scientific Data*, 2016, 3: 160018.
- [44] Gebru T, Morgenstern J, Vecchione B, et al. Datasheets for datasets[J]. *Communications of the ACM*, 2021, 64(12): 86–92.
- [45] Roscher R, Bohn B, Duarte M F, et al. Explainable machine learning for scientific insights and discoveries[J]. *IEEE Access*, 2020, 8: 42200–42216.
- [46] Krenn M, Pollice R, Guo S Y, et al. On scientific understanding with artificial intelligence[J]. *Nature Reviews Physics*, 2022, 4(12): 761–769.
- [47] Psaros A F, Meng X H, Zou Z R, et al. Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons[J]. *Journal of Computational Physics*, 2023, 477: 111902.
- [48] Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science[J]. *Patterns*, 2023, 4(9): 100804.
- [49] Mitchell M, Wu S, Zaldivar A, et al. Model cards for model reporting[C]//Proceedings of the Conference on Fairness, Accountability, and Transparency. New York: ACM, 2019: 220–229.
- [50] Walters W H, Wilder E I. Fabrication and errors in the bibliographic citations generated by ChatGPT[J]. *Scientific Reports*, 2023, 13: 14045.
- [51] Farquhar S, Kossen J, Kuhn L, et al. Detecting hallucinations in large language models using semantic entropy[J]. *Nature*, 2024, 630(8017): 625–630.
- [52] Kalai A T, Nachum O, Vempala S S, et al. Evaluating large language models for accuracy incentivizes hallucinations[J]. *Nature*, 2026, 653(8116): 1047–1051.

Artificial intelligence in astrophysics: Foundation models and scientific agents

WANG Yu^{1,2}, LI Liang³, CAI Rong-Gen^{3*}

1. International Center for Relativistic Astrophysics (ICRA), Rome I-00185, Italy
2. Osservatorio Astronomico d' Abruzzo, National Institute for Astrophysics (INAF), Teramo I-64100, Italy
3. Institute of Fundamental Physics and Quantum Technology, Ningbo University, Ningbo 315211, China

Abstract With the rapid development of wide-field surveys, high-cadence monitoring, and multimessenger observations, astronomy is facing a new era of simultaneous growth in data volume, data variety, and observing frequency. At the same time, artificial intelligence is expanding from task-specific models to foundation models and large-language-model-based scientific agents. This article reviews representative advances of these methods in astrophysics over the past five years, with a focus on comparing their functional boundaries in terms of training objectives, data organization, transfer strategies, and scientific roles. The review shows that task-specific models remain the most mature and effective for well-defined problems with stable data distributions; foundation models, pretrained on large-scale heterogeneous data, have demonstrated promising potential for unified representation and cross-task transfer in imaging, spectroscopy, and time-series tasks; scientific agents, however, are still in early exploratory stages, and their roles in literature summarization, tool invocation, and multi-step research workflow orchestration require cautious assessment. The article further examines key challenges, including data quality, generalization, physical interpretability, uncertainty quantification, and reproducibility. It identifies future priorities in building high-quality data infrastructure, establishing cross-survey out-of-distribution validation standards, solidifying software toolchains, and fostering interdisciplinary talent training.

Keywords astrophysics; artificial intelligence; foundation models; large language models; scientific agents; multimodal learning ●



(责任编辑 傅雪)