DOI: 10.7544/issn1000-1239.qy20260901 CSTR: 32373.14.issn1000-1239.qy20260901

前　言

近年来,以生成式大模型为代表的人工智能技术取得了长足进展,在自然语言处理、多模态理解、逻辑推理、代码生成等领域广泛应用,成为推动科技创新和数字经济的重要基础设施。从 Transformer 架构的提出,到大规模预训练模型在多任务场景中的涌现能力,再到多模态基础模型的加速演进,当前人工智能呈现出“参数规模持续扩张、任务性能不断刷新的快速迭代态势”。然而,伴随“更大模型、更大数据、更大算力”的发展路径,现有人工智能技术的内在局限也日益凸显。首先,当前主流大模型高度依赖超大规模语料,而训练数据中存在的大量冗余、噪声和过期信息削弱了结构性模式的学习效果,难以抽象出高质量、可复用的知识结构与因果机制,整体数据利用效率偏低。正如图灵奖得主 Bengio 等所指出,缺乏结构化知识与因果建模支撑的纯统计学习模式难以实现可靠泛化和系统性推理。其次,大模型虽在开放域任务中表现突出,但在专业知识密集型场景中仍频繁出现事实性错误与逻辑不一致的“幻觉”现象,反映出其内部与显式知识形式尚未有效对齐,推理过程难以被充分刻画、解释与验证。其三,传统知识工程强调知识来源、适用范围与责任主体,而大模型通过端到端训练以隐式方式“吸收”知识,模型缺乏明确知识边界与不确定性表征,会显著放大模型在分布外场景下的失效风险。

在上述背景下,知识增强的人工智能技术范式逐渐成为国际研究热点。其核心目标是旨在保持数据驱动优势的同时,系统引入符号知识、结构化知识、多模态知识、人类认知规律等,使模型在学习、推理和决策过程中能够显式表征和复用知识,实现与人类认知结构的更紧密对齐;通过可追踪的推理链条支撑结果解释与过程验证,提升技术的可验证性与可控性;支持知识的持续更新与按需注入,缓解灾难性遗忘并增强对动态环境的适应能力;围绕这一方向,检索增强生成、知识图谱增强、神经符号计算、人机协同的可验证推理、多模态知识注入等技术路线快速发展,有望成为迈向“可信、可控、可解释”下一代人工智能的重要支柱。

为进一步推动我国学者在人工智能技术与应用研究,汇聚领域内最新研究成果与典型应用案例,《计算机研究与发展》组织策划了“知识增强的人工智能前沿技术及应用”专题。自征稿以来,本专题共收到来稿 62 篇,其中 57 篇通过形式审查进入评审阶段,经初审、复审、第十一届数据挖掘会议 (CCDM2026) 宣读和终审等阶段,最终遴选录用 10 篇

稿件，内容涵盖认知理解、知识学习、智能推理、人机增强、多模态理解与生成、模型安全性、智能教育与智能医疗应用等多个方向，充分展现了当前国内在知识增强人工智能的方法创新与应用实践中的蓬勃活力，为同行学者提供参考与启发。

马博翔等作者的研究“大语言模型的齿状回：基于模式分离的场景认知能力诊断及增强框架”针对大语言模型在场景认知任务上的系统性泛化局限问题，受人脑海马体中齿状回结构的模式分离机制启发，提出跨实体表示诊断方法，在无需额外训练的条件下量化模型各层实体表示的混淆程度。在此基础上，提出隐空间模式分离方法，在混淆较严重的层引入三元组对比正则，将模式分离约束显式融入大模型训练过程。该研究从内部表示可诊断性与知识“去混淆”角度出发，为理解和改进大模型的知识记忆机制提供了新的分析框架。

陈伟等作者的研究“多视图增强图注意力网络的多模态知识图谱补全”提出一种基于多视图增强图注意力网络的多模态知识图谱补全方法，将结构、视觉与文本信息统一建模为共享拓扑下的3个独立视图，为每一视图构建关系感知图注意力编码器，实现关系约束下的邻域传播和结构语义建模，并通过门控后融合机制动态整合多视图表示，有效缓解异构模态直接融合带来的语义干扰。该研究为构建高质量、多模态融合的知识基座提供了有效手段。

王迪等作者的研究“语义对齐与自适应回放增强的动态多模态知识图谱持续补全”针对动态场景下知识持续补全的问题，提出语义对齐与自适应回放增强的动态多模态知识图谱持续补全方法。该方法利用大语言模型生成实体语义描述，结合视觉-语言预训练表征构建多模态差异感知融合机制。通过多模态时序语义对齐约束实体表征在快照间的演化一致性，并通过时序感知自适应回放机制选择性回放历史知识，缓解灾难性遗忘。该研究为开放环境中的持续知识注入与自适应模型能力增强提供了可行路径。

邵杰晶等作者的研究“面向可解释人机协同的知识引导主动学习”提出面向可解释人机协同的知识引导主动学习方法，以一阶逻辑规则作为模型与专家的协同媒介，将模型预测行为归纳为紧凑规则，由专家判别其正确性。对正确规则进行演绎式弱标注扩展，将错误规则视为知识与模型分歧的符号化观测以驱动反绎式分歧查询，从人机协同视角重构传统主动学习范式。该研究验证了多个任务、特别是小预算标注场景中方法的稳定性优势，为将专家知识显式融入主动学习流程、构建可解释的人机协同学习机制提供了范例。

司悦航等作者的研究“面向时序知识图谱推理的本体-图谱双智能体强化学习方法”针对现有强化学习方法依赖实体级路径搜索而缺乏高层语义指导的问题，提出面向时序知识图谱推理的本体-图谱双智能体强化学习方法。利用大模型语义嵌入与时序图谱推理模型结构嵌入联合初始化知识本体，并通过可学习频次权重实现本体动态优化；随后采用双智能体强化学习，上层智能体在本体图上进行概念级游走，为下层智能体在时序图谱

上的实体级搜索提供语义导航与分层奖励约束。该研究验证了本体引导与双智能体架构在时序推理中的有效性，为时序知识推理的强化学习范式提供了有效方法。

马杰等作者的研究“基于多智能体辩论的图文交织生成方法”面向多模态检索增强生成任务，提出一种基于多智能体辩论机制的多模态检索增强图文交织生成方法。该方法将生成过程解耦为文本推理、视觉感知和全局裁决3类协同角色，并通过双轨草稿生成、冲突检测、多轮对抗式辩论及全局裁决与融合机制，实现图像选择、动态补图、冗余剔除与顺序重排的统一优化。该研究验证了基于多智能体协同与辩论机制的知识选择与整合能力，为智能体协同能力提升提供了创新思路。

武栩含等作者的研究“基于常识引导与多模态特征融合的视觉关系推理”提出一种嵌入常识知识的端到端场景图生成模型，针对场景图生成中逻辑幻觉与缺乏语义约束的问题，提出软对齐知识融合机制，将视觉特征与常识知识图谱嵌入空间进行概率映射与动态聚合，并通过常识感知与知识评分约束联合训练，提升视觉关系推理的语义一致性。该研究验证了常识与多模态融合引导的建模方式在低频关系识别和减少语义冲突预测等方面具有优势，为知识增强的多模态视觉关系推理提供重要支撑。

曹明生等作者的研究“基于大语言模型的源代码漏洞检测方法”面向源代码安全场景，提出一种融合图结构表征与大语言模型推理能力的漏洞检测方法。该方法以数据流图为基础建模源代码变量依赖与语义传播路径，通过跨模态映射实现程序结构知识的有效注入。在此基础上，采用基于代码重构的渐进式图编码器预训练、结合结构化推理蒸馏与LoRA等高效微调，在降低训练开销的同时提升漏洞判别性能；并引入面向漏洞检测的思维链推理以增强可解释性与一致性。该研究体现了知识增强用于提升大模型安全检测方面的可信度和实用性潜力。

龙佳琪等作者的研究“CAKT-Agent: 基于MoE的情境自适应知识追踪智能体”针对传统知识追踪方法在情境适配与可解释性上的局限，提出了一种基于混合专家模型的情境自适应知识追踪智能体。通过构建统一记忆模块、学习智能体与伴学智能体三级协同架构，为个性化学习路径规划提供依据。该研究验证了教育领域知识与大模型智能体深度融合驱动领域应用研究的必要性，强化了领域知识增强的人工智能在高利害教育场景下的实践价值。

吕千一等作者的研究“PathRoute: 基于全切片图像的主动证据路由的癌症生存预测方法”面向病理全切片生存预测任务，提出一个面向可解释生存预测的主动证据路由框架。将病理全切片图像上的生存预测过程建模为“证据形成—风险读出”的两阶段有向推理过程，通过在统一隐空间中解耦4类功能明确的可学习查询（组织感知、路由聚合、表型语义、风险读出），并引入方向性注意力掩码，实现了对复杂病理特征的结构化建模与信息流的严格约束。该研究强化了知识增强的人工智能技术在临床预后评估中的有效性与鲁棒

性，为人工智能在医疗场景下的辅助诊疗提供了重要基础。

本专题系统汇聚了知识增强人工智能在基础机制、关键技术与典型应用上的最新进展，表明知识增强范式对提升人工智能大模型的可解释性、可验证性与可控性具有重要意义。面向未来，仍需在统一知识底座构建、跨模态/跨领域对齐、持续学习与安全治理等方向持续攻关，推动知识增强与大模型预训练、推理框架的深度协同与规模化落地。受篇幅限制，本专题无法全面覆盖该领域的所有研究进展，不足之处敬请各位学者谅解和批评指正。感谢《计算机研究与发展》编辑部、专刊编委会和CCF人工智能与模式识别专业委员会的悉心指导。感谢专题全体评审专家及时、耐心、细致的评审工作。感谢所有作者的积极参与与高质量投稿。期望本专题能够抛砖引玉，为学界与产业界提供参考与启发，持续推动知识增强人工智能在理论创新与工程实践中的高质量发展。

尹义龙 (山东大学)
白亮 (山西大学)
王士进 (科大讯飞股份有限公司)
黄振亚 (中国科学技术大学)
2026年8月