前　言

近年来,随着科技和生产力的飞速更新,AI大模型迈入规模应用的新阶段,人工智能成为助推科技高质量发展、赋能千行百业的重要推手.数据作为核心生产要素,已逐渐成为国家重要战略资源.存储作为承载数据的重要载体,已成为释放数据价值的重要基石,数据潜能的有效激发取决于数据的高效分析处理.那么,大模型及其应用的出现给传统的存储管理带来哪些新的挑战?大模型是否会完全取代传统机器学习模型,并在数据分析处理中取得新的突破?大模型时代的存储管理与数据分析已经成为学术界和产业界广泛关注的焦点.

为促进存储领域技术交流,《计算机研究与发展》推出了本期“大模型时代的存储管理与数据分析”专题.本专题收录了6篇论文,聚焦大模型给存储管理与数据分析带来的机遇与挑战,深入探讨支持大模型预训练、微调、评估和推理等的存储管理技术,以及基于大模型的数据分析处理技术,希望能为从事相关工作的读者提供借鉴和帮助.

葛旭冉等作者的论文“大语言模型推理中的存储优化技术综述”研究分析了大语言模型推理系统在存储方面面临的诸多问题,从显存优化、异构存储和分布式存储等方面归纳总结了现有研究工作,从探索新型硬件和边缘设备利用等方面对未来研究方向进行展望.

台建玮等作者的论文“大语言模型对抗性攻击与防御综述”研究分析了近年来针对大语言模型的对抗性攻击方法和防御策略的基本原理、实施方法与研究结论.深入探讨了提示注入攻击、间接提示注入攻击、越狱攻击和后门攻击等主流攻击模式,进一步对大语言模型安全的未来发展方向进行了展望.

尚碧筠等作者的论文“ScaleFS:面向大语言模型的高性能可扩展元数据设计”针对传统存储系统在大语言模型场景下面临的主要瓶颈问题,研究分析了大语言模型在数据准备、模型训练和推理阶段的存储访问特征,设计实现了一种高性能、可扩展的分布式元数据,通过目录树元数据与属性元数据解耦的架构设计,结合深度与广度均衡的目录树分层分区策略设计,提高路径解析、负载均衡和系统扩展能力.

吴文隆等作者的论文“大语言模型和知识图谱协同的跨域异质数据查询框架”针对大语言模型对跨域异质数据的共享利用问题,研究提出了一种大语言模型和知识图谱协同的跨域异质数据查询框架,采用适配器对跨域异质数据进行融合,并构建相应的知识图谱,提出同源知识图抽取算法来实现知识图谱的重构,提高查询性能.提出可信候选子图匹配算法,保证多域数据查询的高可信度.提出基于线性知识图提示的多域数据查询算法,实现LLM+KG范式下的高效可信跨域查询.

赵安宁等作者的论文“面向低磨损存内计算的多状态逻辑门综合”针对传统“冯•诺依曼”计算系统的“存储墙”和“能耗墙”,在研究分析近年来提出的一系列存内状态逻辑门的基础

上，深入研究了复杂计算过程到存内状态逻辑实现的自动化综合映射方法，以降低复杂存内状态逻辑计算过程的磨损（翻转率）为目标，研究提出了一种面向低磨损存内计算的多状态逻辑门综合映射方法.

郝栋栋等作者的论文“面向SCSI子系统的用户空间存储架构设计”针对慢速存储设备软件栈的时间开销问题，设计了硬件性能测试工具HwPerfIO，能够更准确地测试低延迟I/O设备的性能，结合测试结果分析了内核I/O栈对性能测试所带来的影响.进一步基于Intel提出的高性能存储开发包SPDK，设计实现了用户空间SCSI存储I/O栈，减少了应用程序与存储设备I/O路径上的软件开销.

由于稿源丰富和专题篇幅有限等原因，本专题无法全面覆盖存储领域各方面的最新研究进展，不当之处请同行学者批评指正！感谢各位作者、审稿专家和编辑部的全力支持和辛勤付出！

舒继武（清华大学）
王意洁（国防科技大学）
2025年2月