

教育大模型数据治理与共享标准框架构建研究

吴永和¹ 陈圆圆¹ 吴慧娜¹ 马晓玲²

(1. 华东师范大学教育信息技术学系, 上海 200062; 2. 华东师范大学信息管理学系, 上海 200062)

[摘要] 智能范式背景下, 教育数据呈现来源异构、授权不清、依赖情境、高敏感性等特征, 带来协作治理、隐私保护、跨域共享等挑战。本研究以高质量教育数据为核心, 提出“安全—质量—语境”的内涵特性, 明确教育数据规范管理与高效利用的核心目标与价值导向; 构建了教育大模型数据治理和共享标准框架, 并从释放数据价值、规范生成数据、强化知识支撑和构建协同机制四方面展望其未来发展。这一框架有助于为教育大模型的预训练和测评提供数据支撑, 促进其安全、高效、可信地应用。

[关键词] 教育大模型; 数据治理; 数据共享; 高质量教育数据; 标准框架

[中图分类号] G434

[文献标识码] A

[文章编号] 1007-2179(2026)01-0054-13

一、引言

随着人工智能技术的迅猛发展及其广泛应用, 以人机共生、跨域融合、智能驱动为特征(周代数等, 2024; 颜世健等, 2024; 罗威等, 2020)的“智能科学”第五范式(AI for Science)应运而生。人工智能大模型作为关键支撑力量, 正展现出重塑教育模式的巨大潜力。2025年1月, 中共中央、国务院印发《教育强国建设规划纲要(2024—2035年)》, 明确提出打造教育人工智能大模型(简称“教育大模型”), 建立基于大数据和人工智能的教育评价和科学决策制度。人工智能大模型的性能与效能离不开高质量、大规模、多样化的数据支持

(Kramcsák, 2023)。然而, 当前教育数据存在的质量参差不齐、来源分散、标准不统一等问题(徐峰等, 2018), 不仅制约着教育大模型性能的提升, 还可能导致模型输出不稳定、偏差加剧, 影响教育应用的公平性与可靠性。数据治理在确保数据质量、提升数据可用性方面发挥着关键作用。数据共享与数据治理相伴相生, 有利于打破信息孤岛, 实现资源优化配置, 为教育大模型提供有力保障, 推动智慧教育的创新与进步。有研究提出多种数据治理与共享的路径和方案(李森, 2023; 马治国等, 2024; 董伟玮等, 2025), 但多侧重于风险防控与制度设计, 对如何保障和提升数据质量关注不足, 且缺乏统一的治理共享框架, 难以满足大模型对高质

[收稿日期] 2025-02-17 **[修回日期]** 2025-08-25 **[DOI编码]** 10.13966/j.cnki.kfjyyj.2026.01.007

[基金项目] 2021年度国家社会科学基金重大项目“面向未成年人的人工智能技术规范研究”(21&ZD328), 2023年度科技部“社会治理与智慧社会科技支撑”重点专项“教育大数据驱动的个性化学习关键技术研究与应用”(2023YFC3341200)。

[作者简介] 吴永和(通讯作者), 博士, 研究员, 博士生导师, 教育部教育信息化技术标准委员会主任委员, 华东师范大学教育信息技术学系, 研究方向: 教育数字化转型、智能驱动教育、模式驱动教育、数字教育技术标准(yhwu@deit.ecnu.edu.cn); 陈圆圆, 博士研究生, 华东师范大学教育信息技术学系, 研究方向: 人工智能教育应用; 吴慧娜, 博士研究生, 华东师范大学教育信息技术学系, 研究方向: 智能教育与学习分析; 马晓玲, 副教授, 华东师范大学信息管理学系, 研究方向: 信息管理、人工智能应用。

[引用信息] 吴永和, 陈圆圆, 吴慧娜, 马晓玲(2026). 教育大模型数据治理与共享标准框架构建研究[J]. 开放教育研究, 32(1): 54-66.

量数据的迫切需求。本研究在深入分析教育数据特性的基础上,构建教育大模型数据治理与共享标准框架,为教育大模型发展提供保障。

二、现实需求

(一)概念内涵

1. 数据治理

学术界和产业界对数据治理有着丰富的理论阐释。从操作层面而言,数据治理指对数据进行处置、格式化和规范化的过程(全国信息技术标准化技术委员会,2017),以满足外部监管需求(Cheong & Chang, 2007)、提高数据质量(Madnick et al., 2009)等。从治理权责角度看,国际数据治理研究所(Data Governance Institute, 2019)将其定义为与数据信息相关的决策与权责体系。国际数据管理协会视其为行使权利和控制数据资产管理的活动集合(Chowdry et al., 2013)。李青等(2018)融合“活动集合”和“权责分配”概念,认为数据治理是数据管理的综合实践。目前教育数据治理主要集中在高等教育领域,其核心在于高校通过系统化的制度、流程和方法,对教育数据进行规范、控制和协调(董晓辉, 2021)。

随着生成式人工智能等新技术在教育领域加速渗透,传统治理模式难以满足教育大模型对数据类型、质量、安全、伦理与隐私的要求。因此,本研究将教育大模型数据治理界定为:围绕用于预训练、微调与评估教育大模型各类教育数据资源,在数据采集、管理、使用和维护等过程中实施的规范化、制度化治理活动。治理核心在于提升数据质量,强化隐私保护,推动教育大模型的科学、公平与可持续发展。

2. 数据共享

共享是实现和提升数据应用价值的基础(田贤鹏, 2020)。数据共享指不同主体之间对数据资源进行有序、合法、可控地交换与利用,其核心在于打破数据孤岛,实现数据的流通与协同价值最大化(王利明, 2019)。在不同应用领域,数据共享有特定的内涵特征。政务服务数据共享指政府部门间通过规范机制实现业务数据的按需流通,以提升政务服务效率(徐晓林等, 2018)。教育领域数据共享主要集中在教育资源的共建共享上,旨在推动教学

资源的互通和增值(孙旭东等, 2024; 钱冬明等, 2013)。与资源共享强调内容供给与获取不同,教育数据共享关注多源数据的融合与协同(张绍丽等, 2017)。教育数据的全过程共享指在教育自然情境中采集数据,以问题和需求为导向,辅助教育科学管理与决策(陈丽等, 2024)。

教育大模型数据共享是数据共享理念在人工智能时代教育领域的深化与拓展,其内涵体现为:在符合伦理规范与数据安全要求的前提下,围绕教育大模型研发与应用,系统实现高质量、多维度教育数据在各类主体之间的协同流通与开放利用。

3. 治理与共享的关系

数据治理与数据共享既相互依存又彼此制约。数据治理旨在确保数据资产得到正确有效的管理(包冬梅等, 2015);数据共享则强调多主体间数据的可控流通与有效利用,以最大化其价值(王利明, 2019)。从逻辑关系看,治理是共享的前提与保障,高质量、可信任的数据是有效共享的基础;共享则是治理成果的延伸与放大,推动数据在多场景中产生复合价值,并反向促进治理机制的优化。

智能范式下两者关系更趋紧密:一方面,教育大模型高度依赖训练数据,严格的数据治理手段是保障数据质量、支撑模型能力演化的根本前提;另一方面,教育数据普遍存在分布分散等问题(周东岱等, 2019),若无法实现跨机构、跨平台的数据共享,将难以提升大模型的泛化能力。因此,在治理保障下推进数据共享,不仅可以提升教育大模型的适应性和鲁棒性,也能最大化利用优质教育数据的价值,服务教学诊断、个性化学习推荐等应用场景。由此可见,治理与共享协同推进,构成了高质量数据资源供给与有效利用的双重基础。

(二)面临挑战

随着人工智能技术的持续演进,数据经历了从静态向动态、多模态的形态转变。不同科研范式下数据的特征差异,深刻影响着数据治理与共享的方式。尤其在当前以大模型为代表的智能范式下,数据规模激增、语义复杂化与生成数据的广泛应用,带来一系列新的治理与共享难题。本研究从不同范式演进的角度出发,系统梳理数据特征的变化(见表1),并揭示智能范式下教育数据治理与共享面临的新挑战。

表 1 不同范式的数据特点

科研范式	技术	数据作用	数据特点	局限	数据治理与共享
第一范式 经验科学	望远镜、显微镜等基础观测工具	辅助	低结构化, 分散, 多为个体经验数据	量小、主观性强	治理无统一标准, 数据不受约束
第二范式 理论科学	微积分等数学工具	验证理论模型	结构化程度提高, 以数学和实验数据为主	数据量有限、获取成本高、依赖理论假设	有一定的科研规范, 数据共享仍有限, 主要由实验室管理
第三范式 计算科学	电子计算机	支持模拟计算和优化	大规模、结构化数据, 多为计算材料和技术数据	缺乏真实数据	数据治理初步成型, 涉及访问控制、元数据管理, 但数据共享仍限于机构内部
第四范式 数据科学	大数据技术、机器学习算法	数据驱动发现	超大规模、多源异构数据	数据噪声大、质量参差不齐、可解释性差	数据开放和共享加速, 科学数据管理原则得到推广
第五范式 智能科学	生成式人工智能、量子计算等	与算法、算力深度融合, 驱动模型自动演进	非结构化、多模态数据	算法偏见、数据伦理安全问题、数据主体边界模糊	数据治理标准化, 强调多方协作治理; 注重数据质量、隐私保护与安全伦理治理

随着科学范式的演进, 数据经历了从“被动记录”到“自主生产力”的转变。智能范式下, 数据成为科学发现的核心主体(张越等, 2025)。然而, 数据的碎片化、简化处理及隐私偏见等问题仍然存在, 并对模型性能与伦理安全构成挑战。智能范式强调多模态数据与人机融合(周代数等, 2024), 对数据的真实性、完整性与可控性提出更高要求。数据质量不足易导致模型幻觉、偏见传播、决策误导等问题。此外, 数据使用主体的快速扩展也加剧了数据权责关系的复杂性(梅傲等, 2024)。在以往相对封闭的数据环境中, 数据的采集、使用与管理往往归属于单一机构或平台, 治理关系清晰。目前数据由用户、平台运营者、算法开发商及第三方服务商等多元主体共同生成、流通与处理, 原有的权属界限逐渐模糊。

教育数据作为驱动教育智能化发展的核心资源, 具有显著的独特性。在教育活动中, 数据不仅来源复杂、类型多样(彭振宇等, 2025), 还深度嵌入教、学、评各环节, 体现出高度的动态性、情境性与个体差异性。智能范式下教育数据继承了传统数据普遍存在的来源分散、授权不清、隐私风险突出等治理与共享难题, 面临更为错综复杂的挑战: 一是教育数据高度依赖具体教学情境(刘三女牙等, 2017), 包含学生行为轨迹、课堂互动、评估反馈等多维度信息, 若数据共享脱离语境或被泛化使用, 易引发语义偏差与决策失真, 影响数据可用性与公正性。二是教育数据敏感性高(唐颖等, 2025), 尤其涉及大量未成年人的学习、心理、社交等信息,

传统监管机制难以有效覆盖其全生命周期。三是共享基础薄弱, 教育数据来源异构、标准不一, 且缺乏统一的数据分类标准和数据格式(赵润娣, 2021), 严重制约数据互联互通和模型协同优化的可行性。因此, 构建以多元协同和多方共识为基础的治理与共享标准, 强化数据隐私保护与伦理风险防控, 已成为当务之急。

(三) 现状与不足

1. 数据治理、共享与标准的关系

数据标准化是数据治理的先决条件与基础环节(程娅, 2022; 李青等, 2018), 也是破除数据管理困境、提高数据质量、释放数据价值的关键所在(中国通信标准化协会, 2021)。在教育领域, 数据标准为教育大数据的智慧管理与科学应用奠定了基础, 为数据资产的质量监控与价值评估提供了量化依据(余鹏等, 2018), 促使来自不同平台、不同地区的教育数据在满足隐私保护与安全合规的前提下高效整合与共享, 从而为大模型的预训练、微调和评估提供稳定、可信且高质量的数据支撑。标准不仅是实现技术互操作性的关键, 更是实现教育大模型“管住数据”“共享数据”“用好数据”的制度基础。

2. 数据治理相关标准

国际层面, 数据治理标准化起步较早且相对成熟, 主流数据治理框架包括国际数据管理协会的数据管理框架(DAMA International, 2017)和国际数据治理研究所的数据治理框架(Data Governance Institute, 2019)。国际化标准组织围绕数据治理提

出了完整的方法论和管理、分类等规范(International Organization for Standardization, 2015; International Organization for Standardization, 2017; International Organization for Standardization, 2022), 欧盟《通用数据保护条例》则为全球数据治理的合规性提供重要参考(Regulation, 2018)。

我国《信息技术服务治理第5部分:数据治理规范》(国家标准委, 2018a)提出了数据治理的总则和框架,《信息技术 大数据 数据治理实施指南》(国家标准委, 2024b)定义了数据治理实施的基本要求及过程。2021年,《数据管理能力成熟度评估模型》(国家标准委, 2018b)正式实施,为数据治理能力评估提供了标准化工具。山东省、哈尔滨市等制定了地方数据标准(山东省大数据局, 2023; 哈尔滨市大数据中心, 2024),为公共数据治理提供支撑。此外,医疗、金融、电信等数据密集型行业也逐步形成针对质量、安全与合规的行业标准。

针对教育大模型的数据治理标准研制相对滞后,但已有进展。2023年世界人工智能大会发布的《教育通用人工智能大模型 第3部分:数据规范》为研究教育大模型数据治理与共享提供了依据(吴永和等, 2024)。科大讯飞与中国电子技术标准化研究院联合发布国际标准《人工智能—分析和机器学习的数据质量—第4部分:数据质量过程框架》(International Organization for Standardization, 2024),覆盖数据获取、准备、标注、评估和使用等流程,为大模型数据治理提供基础指导。

3. 共享相关标准

为推动数据的安全、合法与高效共享,各国相继出台相关标准来规范数据交换和共享行为(杨现民等, 2023)。国际层面,欧盟通过颁布系列战略引导成员国开放共享数据(周力虹等, 2025)。2011年,欧盟颁布《开放数据战略》,提出“泛欧门户”,推动无障碍信息共享;2013年,各成员国共同签署《数据开放宪章》,明确将教育数据列为优先开放的数据类别;2020年发布的《欧洲数据战略》强调加强数字基础设施建设,提升数据存储、处理、使用和互操作能力(European Commission, 2025)。英美等国侧重于操作性指导,如英国信息专员办公室发布《数据共享业务守则》,要求以公平、安全和透明的方式共享数据,美国通过《数据目录快速指南》

《数据目录终极指南》等提出二十个标准,界定数据目录的内涵、主体、功能和步骤等,形成支持数据共享的组织实施框架(Centers for Medicare& Medicaid Services, 2022; Wells, 2018)。

我国《国家数据标准体系建设指南》(国家发改委等, 2024)指出,数据共享标准主要规范程序要求、系统平台、评价要求等。《信息技术 大数据 政务数据开放共享》系列标准(国家标准委, 2020),细化了开放共享的网络、数据、平台、安全要求。行业标准层面,交通运输、金融、互联网、农业、气象等行业出台了数据共享标准,规定业务数据共享指标的分类、描述方法等(中国电子技术标准化研究院, 2024)。地方标准层面,上海市、山东省构建了数据共享地方标准(上海市大数据中心, 2020; 山东省教育信息中心, 2019)。

4. 主要不足

当前数据治理和数据共享的通用标准体系已初具规模,但仍面临挑战:一是标准整体呈现碎片化倾向,缺乏统一的顶层设计,各环节衔接不畅;二是新兴领域标准供给不足,缺乏针对多模态、隐私敏感数据的数据治理与共享标准,难以满足实际应用场景的技术需求与合规要求;三是现有标准聚焦数据格式、接口和安全等属性,缺乏面向“高质量数据”的指标体系,难以有效支撑教育大模型对数据完整性、可用性和可信度的需求。

三、理论框架

(一)教育数据治理与共享的新要求

在科学范式不断演进的背景下,教育数据治理与共享亟需适应新的技术条件、应用场景和伦理要求。教育数据治理与共享亟需以“高质量数据”为导向,从“量的积累”转向“质的跃迁”。

1. 强化高质量数据导向的治理目标

教育大模型的发展与教育数据紧密相关,高质量数据不仅是支撑大模型预训练与微调的基础,更是保障其理解教育情境、提供个性化支持、生成教育内容的关键。教育数据长期存在内容杂(刘洋溪, 2024)、准确性不足、标准不统一(徐峰等, 2018)等问题。面对智能范式下模型能力演化的迫切需求,本研究认为有必要强化以高质量数据为导向,系统审视“高质量教育数据”的内涵与特征。

2. 加强数据安全与伦理隐私治理

数据安全与伦理是教育数据广泛共享与有效利用的前提。教育数据中包含未成年人的学习行为、心理状态与社交互动等敏感信息, 处理不当可能侵害师生合法权益, 削弱公众信任(唐颖等, 2025)。特别是大模型深度参与教育应用的背景下, 算法决策的不透明性与数据处理的不可控性会放大伦理风险。这就需要加强数据处理的合法性与正当性, 推动形成教育场景可遵循、可执行的隐私保护规范体系。

3. 推动多元协同的治理机制建设

教育数据来源高度异构、体量庞大、内容繁杂, 传统分散治理的范式存在效率低下、数据滥用、流通和运用乏力等问题(刘洋溪, 2024), 严重制约了教育数据的系统性共享与模型协同优化能力。因此, 智能范式下的治理体系需由单一主体主导向多元协同共治转型, 标准框架需明确各类主体在数据采集、管理、使用与监督等环节的责任边界与协作模式, 并通过信任机制、审查制度与绩效评估等提升治理效能。

4. 完善教育领域的的数据分类体系

目前教育数据在分类体系、格式标准、共享接口等方面仍不统一(苏福根等, 2025), 数据孤岛现象严重, 限制了跨平台、跨场景的高效流通。针对这一问题, 标准框架需提供统一的数据分类规则、元数据标准与共享接口模板, 构建覆盖数据采集、标注、传输、交换全过程的共享技术规范体系, 提升教育数据的互操作性与复用价值, 为教育人工智能的发展提供结构化、可持续的资源保障。

综上, 智能范式下的教育数据治理与共享不仅是对技术标准的更新, 更是对治理理念、价值导向和制度体系的重构。构建以高质量、安全性、协同性为核心导向的标准框架, 是推动教育大模型规范化、可持续发展的重要路径。

(二) 教育数据治理与共享理论框架

为更好地支撑教育大模型的预训练与测评, 构建科学、系统的数据治理与共享理论框架尤为重要。智能范式下教育数据治理与共享理论框架(见图1)以“高质量数据”为核心目标, 围绕治理与共享展开, 并以标准作为互联互通和制度落地的关键手段。四大要素逻辑关联、功能互补, 为标准框

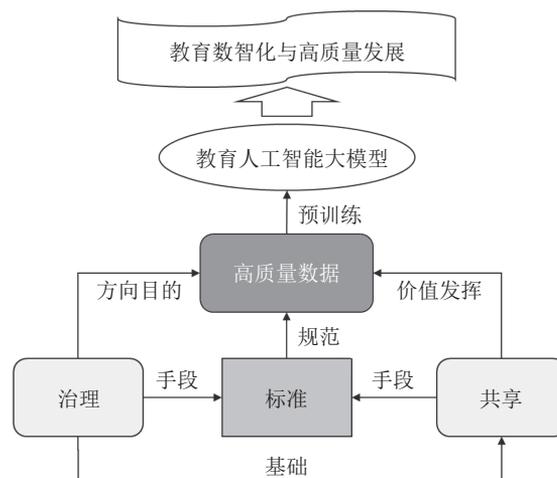


图1 教育大模型数据治理与共享理论框架

架的具体设计奠定理论基础。

1. 高质量数据: 治理与共享的核心目标

高质量教育数据是推动教育人工智能模型实现科学训练、精准决策与可信可用的前提。它应结构清晰、语义一致、时效精准、安全可信等, 还能准确表达教学行为与学习语境, 支持大模型在多场景的泛化能力与可解释性。本研究以高质量数据的内涵特性为起点, 推进贯穿数据治理与共享全过程的标准化设计。

2. 数据治理: 数据管理与安全保障机制

数据治理是确保教育数据合法合规、安全可信的基础机制。教育数据治理覆盖数据生命周期全过程, 强调多主体参与和协同管理。治理体系需兼顾技术性与规范性, 在明确权责的基础上, 加强数据脱敏处理, 并完善伦理审查与风险控制机制, 确保数据合法、合规与安全。同时, 科学合理的数据分类制度能为差异化治理策略的制定与执行提供依据, 提升数据治理的精细化水平与可操作性。

3. 数据共享: 促进数据价值释放的路径

教育数据共享的本质在于推动数据资源在更广泛、更智能的场景中实现价值最大化。面向教育大模型的共享机制应以“受控开放、安全可追溯”为前提, 构建统一的数据共享平台和访问机制, 支持大模型的预训练、微调、测评与持续优化, 加速大模型与教育业务的深度融合。

4. 数据标准: 规范治理与互联互通的基础

标准是实现数据规范治理与可信共享的制度根基。教育数据的多源异构性决定其在处理、集

成与共享中的复杂性和多样性,系统的标准框架可减少语义壁垒与格式差异,提升数据的互操作性。标准的建立可以统一数据格式、分类体系和描述规范,有助于为教育大模型提供结构清晰、语义一致、可复用的数据资源。

四、框架构建

(一)数据为基:高质量教育数据的内涵与特征

1. 内涵

学者们从不同角度界定了教育数据。希尔德坎普等(Schildkamp et al., 2012)指出,教育数据指通过定性或定量的方法收集的学生、教师、领导和学校等的信息。谢娟(2020)将教育数据界定为与教育过程中个体相关的信息或数据产物。杨现民等(2017)认为教育数据广义上源自教育活动中的人类行为数据,狭义上来源于教育管理系统和在线平台的学习者行为数据。

高质量的教育数据应始终“以人为本”,关注学生的全面发展和个性化需求,确保教育的社会责任和育人目标不被忽视。在教育数据的应用过程中,数据安全与伦理问题同样至关重要。构建有效的“数据栅栏”,形成严格的数据隐私保护机制和伦理规范是确保教育数据安全使用的前提(唐颖等, 2025)。高质量教育数据指在真实教育情境中采集的能全面、准确、规范地表征教育活动主体(如教师、学生)、客体(如课程、资源)及其行为与状态的数据集合。这类数据能有效支持教育大模型的训练与测评,保障智能系统运行的科学性、公平性与可靠性,是实现教育智能化发展的基础。

2. 高质量教育数据的特征

美国教育数据治理委员会提出五条判断教育数据质量的标准:及时性、完整性、准确性、有效性和可用性(EDFacts Data Governance Board, 2018),为理解“何为高质量教育数据”提供了参考。结合教育场景的特殊性和大模型的数据依赖性,本研究将高质量教育数据分为三个维度:数据安全、数据质量与数据语境(见图2)。数据安全为教育数据的合法合规使用提供制度保障。数据质量强调准确性与可信性,支撑教育模型训练与结果解释。数据语境关注教育数据的内在属性与外在表现,为教育数据的深度挖掘与广泛应用提供可能。

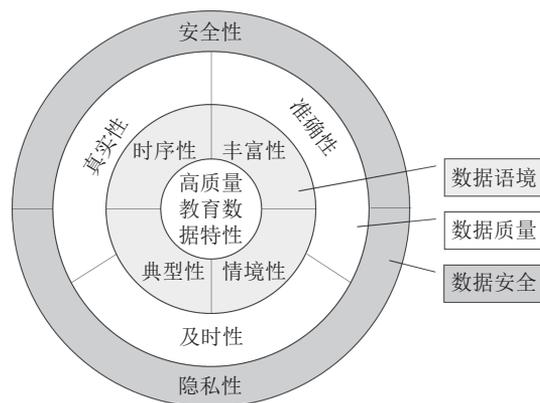


图2 高质量教育数据特性

(二)治理为要:标准框架的核心要素

标准框架是指导数据治理活动开展、协调多方参与主体行为、规范数据资源管理的根本依据。董晓辉(2021)基于治理基本要素(丁志刚, 2014)和活动理论提出高校教育数据治理体系要素框架,包括治理主体、治理客体、治理目标和治理方式。刘洋溪(2024)从敏捷治理角度提出治理目标、治理主体、治理过程、治理工具四维度框架。《数据治理标准化白皮书》提出数据治理“4W1H”模型,聚焦治理的主体、目的、对象、手段与过程。结合已有数据治理理论和相关标准,本研究提出教育大模型数据治理标准框架,它由五个核心要素构成:治理目标、治理主体、治理对象、治理过程与治理内容(见图3)。

1. 治理目标

治理目标是开展各类治理活动的根本导向。

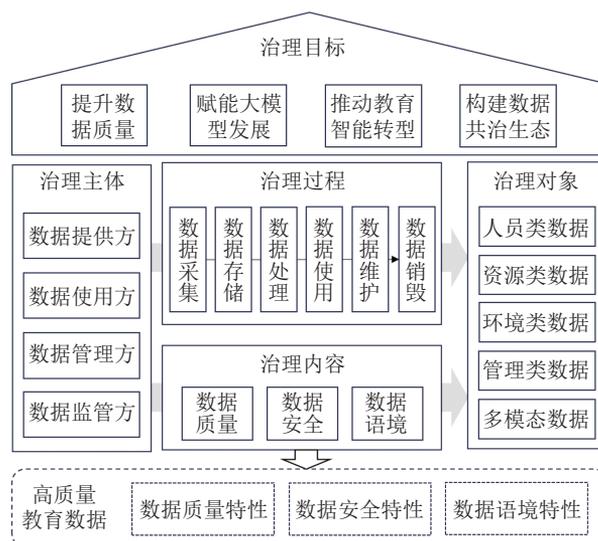


图3 教育大模型数据治理标准框架

它包括三个层次:微观层面侧重于提升教育数据质量与安全水平,为教育大模型构建稳固、可信的底层数据基础;中观层面强调构建数据共享与生态协同机制,推动多主体、多场景的数据融合与价值释放,同时强化高质量数据对模型预训练、微调与迁移的支撑,增强大模型的演进能力;宏观层面以数据治理为支点,推动教育大模型在个性化教学、智能评估、教育治理与教师赋能等关键场景的深度学习应用,重构教育供给方式与学习范式,推动教育数智化与高质量发展。

2. 治理主体

治理主体即从事治理活动的行为者。数据治理不仅仅是技术问题,还涉及多方利益相关者的参与和合作(董晓辉,2021)。智能范式背景下,数据的网络化分布与技术的高度复杂性打破了传统单一机构主导的治理模式,催生出跨领域、跨层级、跨组织的协同治理需求(陈桂香等,2023)。刘洋溪(2024)将教育数据治理主体分为决策层、管理协调层、执行层。张欣(2023)认为在大模型的产业链中至少有四类行动主体:开发者、部署者、用户和接受者。本研究将教育大模型的数据治理主体分为四类(见表2)。

3. 治理对象

治理对象界定了数据治理工作的作用对象。高校教育数据治理的对象包括高校教育数据及相关事务(董晓辉,2021)。本研究的治理对象指大模型预训练、微调、测评等涉及的各类教育数据资源,包括但不限于教学数据、学习行为数据、教育管理数据等。鉴于教育领域数据的复杂性和多样性,本研究建议建立科学的分类体系,以便提高数据治理实践的针对性和高效性。

教育数据有多种分类方式,如分数据业务来源、结构化程度和产生环节三个维度(杨现民等,2016),分不同层级的主体和教育教学活动内容(孙洪涛

等,2016),以及从数据持有者视角和数据功能角度分类(胡新瑞,2025)等。鉴于教育领域尚缺乏专门、统一的数据分类分级标准,杨伟平等(2025)基于《数据安全技术 数据分类分级规则》(国家标准委,2024a),从数据主体、业务领域维度提出适用于教育领域的分类规则。

然而,教育大模型对数据的规模、质量、多样性提出了更高要求,需提出更有针对性的数据分类体系。本研究从教育活动的主体、客体、载体、过程四个维度以及数据的多模态技术特征出发,将教育大模型数据分为人员类数据、资源类数据、环境类数据、管理类数据和多模态数据五类。

人员类数据包括以人口学为代表的基本信息,如学生数据、教师数据和管理者数据,为大模型构建教育主体认知模型提供支撑;资源类数据涵盖教育过程中创建或累积的教学和研究资源,包括教育内容资源、评测练习资源、学习辅助资源和实践学习资源,为大模型提供知识内容和教育语料(Brown et al., 2020);环境类数据指设施设备、平台存储的物理环境数据和虚拟环境数据;管理类数据指涉及管理、行政、决策及统计的数据,支持大模型在教育行政、资源配置等的智能决策;多模态数据指通过多种感知方式收集的与教育活动相关的数据,包括师生的生理数据、心理数据和行为数据(李新等,2024)。上述各类数据的协同为教育大模型的预训练和评测提供支撑。

4. 治理过程

数据生命周期指数据从采集、获取到销毁的全过程,是一切数据活动的基础(张绍华,2016)。教育大模型数据治理本质上是围绕数据全生命周期展开的系统性治理活动,涵盖数据采集、存储、处理、使用、维护和销毁等环节。这一过程应坚持全过程治理理念,确保每一阶段均符合数据安全与合规性要求。

表2 教育大模型数据治理主体

治理主体	角色举例	职责与作用
数据提供方	教师、学生、教育机构、家长和教育内容创作者等	生成、收集并向数据管理方等提供教育数据,数据授权
数据使用方	大模型开发者、教师、学生、研究机构、决策者等	基于授权数据进行分析、训练,开展研究
数据管理方	教育机构的内部管理部门、第三方数据管理机构等	负责数据的采集、存储、清洗、标注等,实施数据脱敏、匿名化
数据监管方	教育部、网信办、信息安全监管机构等	负责制定、执行和监督教育数据使用规则和法规,确保教育数据的使用符合伦理、法律和社会规范,并保障数据的安全性、隐私性及合规性

5. 治理内容

治理内容体现标准框架中需要被明确规范的具体维度,是技术手段与制度要求的结合点。本研究构建了系统性治理矩阵(见表3)。其中,治理内容维度基于高质量数据的特性要求,涉及数据质量治理、安全治理和语境治理等。

表3 数据治理矩阵

生命周期	治理内容			治理主体
	数据质量	数据安全	数据语境	
数据采集	来源可靠 数据完整 实时采集	知情同意 信息加密 来源可追溯	多元化采集 多维度采集 真实场景采集	数据提供方、 数据管理方、 数据监管方
数据存储	完整性维护	加密存储 访问权限	存储可扩展	数据使用方、 数据管理方、 数据监管方
数据处理	一致性管理 数据质量评估	过滤不良信息 标注统一 脱敏处理	数据整合 数据规范化	数据使用方、 数据管理方、 数据监管方
数据使用	准确性度量	隐私保护 避免滥用	趋势分析 周期性分析 典型数据筛选	数据使用方、 数据管理方、 数据监管方
数据维护	数据更新	防止泄露	×	数据管理方、 数据监管方
数据销毁	×	防止泄露	×	数据管理方、 数据监管方

数据质量治理应确保数据的真实性、准确性和及时性。真实性要求数据完整、未被篡改或丢失,不同来源的数据保持一致,避免逻辑冲突;准确性强调数据能真实反映教育场景与个体特征,其保障可通过与权威数据源比对、质量评估等方式实现;及时性要求数据反映当前状态,有赖于更新机制、实时采集技术和时效性监控体系。

数据安全治理旨在确保数据来源安全,防止未经授权访问、泄露或篡改教育数据。这包括对数据来源进行安全审查与溯源管理,确保内容合法合规,避免非法信息干扰;采集前应取得数据主体的知情同意,注重保护个人隐私;在数据标注环节,数据管理方需制定统一标注规范,开展标注质量审核,提升数据标注的安全性与准确性。

数据特征治理强调教育数据的代表性、丰富性与情境适配能力。数据采集手段应多元化,覆盖教育教学的关键环节与场景;数据筛选应注重典型

性,使数据具备推广意义;时间序列数据分析应注重周期性、趋势性,为教育大模型的动态预测与适应性学习提供支持。

(三)共享为径:数据共享的标准化框架

《信息技术 大数据 政务数据开放共享》国家标准基于网络设施、数据资源、平台设施、安全保障和管理评价建立了开放共享系统参考框架(国家标准委,2020)。本研究提出教育大模型数据共享标准框架,由五个模块构成:共享主体、共享数据、共享平台、共享流程、安全与伦理隐私(见图4)。

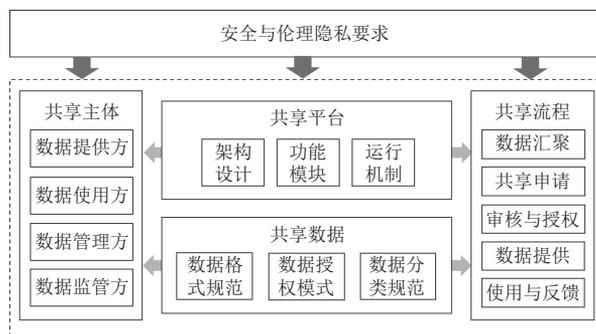


图4 教育大模型数据共享标准框架

1. 共享主体

教育大模型数据共享涉及多方利益相关者,且各方参与需求呈多元化特征。对于公共数据而言,关键参与主体有数据提供方、数据需求方、授权运营方、开发利用方、平台运营方、监督管理方(华为技术有限公司等,2023)。王娟等(2022a)指出,教育政务数据开放共享是政府、企业、学校、社会、公众等主体协同参与的结果。教育大模型的数据共享主体涵盖多层级、多类型的参与方,其分类逻辑应与数据治理体系一致,主要包括数据提供方、数据管理方、数据使用方和数据监管方(见表4)。不同主体的职责应通过制度予以明确,实现权责清晰、协作有序。

2. 共享数据

《电子政务标准化指南 第4部分:信息共享》(国家标准委,2017)规定了信息共享应考虑的数据内容、数据编码、数据质量和数据表达。鉴于教育大模型在训练和应用中对数据完整性、合法性和多样性的高度依赖,数据共享标准应重点关注数据的技术规范、权限管理和内容治理。

数据格式规范主要解决“技术互通”问题,统

表 4 教育大模型数据共享主体

共享主体	角色举例	职责与作用
数据提供方	各级各类教育机构、教育管理部门、教育技术企业等	提供高质量的教育数据资源,是数据共享生态基础
数据管理方	数据服务公司、第三方平台运营机构等	建设维护共享平台,建立数据目录检索系统等
数据使用方	人工智能大模型研发机构、高校科研团队、教育软件开发商等	明确数据使用目的,合规申请与使用共享数据,反馈数据使用成效与问题等
数据监管方	各级教育主管部门、网信、市监等部门	监督数据共享合规性与伦理性,数据安全与隐私保护审查

一格式标准包括元数据标准、数据结构、数据交换格式及编码方式,确保有效传输、解析和融合不同教育机构和系统的数据。数据授权模式主要解决“权限管控”问题,数据管理方可基于数据敏感性、使用目的和主体资质构建分级授权体系,制定标准化的使用协议模板、权限申请审批流程,确保数据有序共享。数据分类规范主要解决“内容治理”问题,即制定不同类别数据的共享策略、安全要求和使用限制,实现对海量异构教育数据的精细化管理和安全化共享。

3. 共享流程

参考山东省大数据局(2024)和国家矿山安全监察局(2023)的共享服务流程,以及王娟等(2022b)教育政务数据开放共享的监管思路,本研究提出面向教育大模型的数据共享流程(见图 5),包括数据汇聚、共享申请、共享审核与数据授权、数据提供、数据使用与反馈五个核心环节。

4. 共享平台

共享平台标准主要规范教育大模型数据共享

平台的架构设计、功能模块与运行机制。平台应具备数据注册、目录管理、权限控制、日志记录、访问审计等核心能力(杨现民等, 2018),支持多角色协同、访问权限设定与全过程监管。平台建设应遵循服务导向与可扩展原则,支持 API 调用、数据可视化展示等功能,并与数据质量评估工具、合规审查工具联动;采用模块化架构,以便针对不同教育场景快速部署和升级平台,推动区域教育数据资源的汇聚共享与智能服务的协同创新(杨现民等, 2024)。

5. 安全与伦理隐私

教育大模型数据共享的安全、伦理隐私标准应以“合法合规、安全可控、责任明确”为核心理念,为数据流通与模型优化提供制度保障,维护教育数据生态的可持续与可信性。这包括:确立数据共享阶段的安全控制机制,如控制访问权限、采用加密技术和加强风险监测,防止数据泄露、滥用及篡改;建立面向教育场景的数据匿名化、脱敏处理标准,确保个人身份不可追溯,特别是在涉及未成

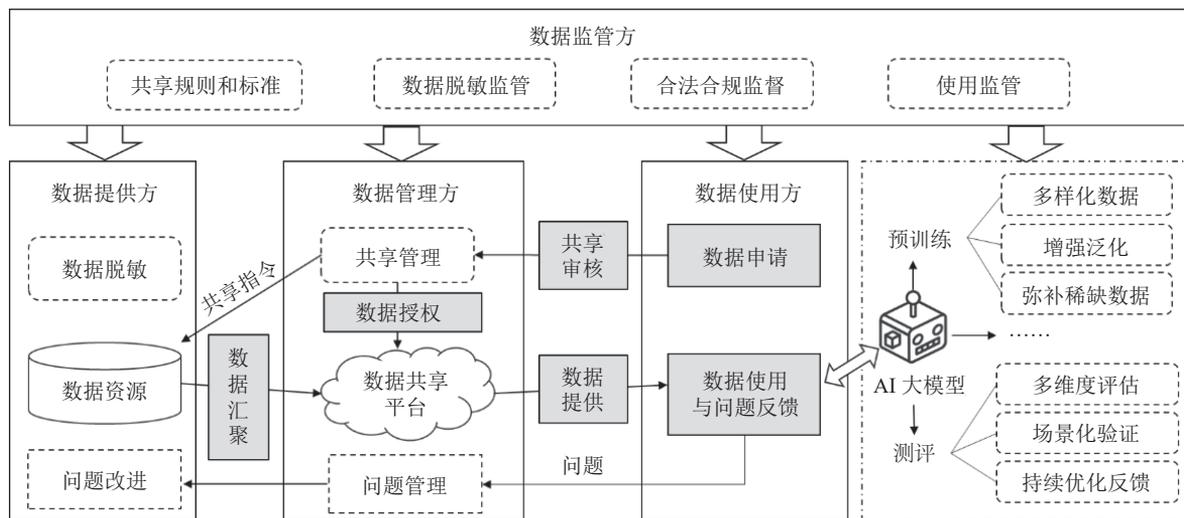


图 5 数据共享流程

年学生数据时需严格落实知情同意机制。

伦理与隐私标准还需回应算法透明性与公平性等问题。数据共享平台应设置伦理审查流程,对涉及敏感人群或高风险用途的数据进行前置审查与动态监管;数据使用方应遵循教育公平公正原则,防范算法偏见对学习产生不当影响。《欧盟人工智能法案》(Council of the European Union, 2024)提出,高风险人工智能系统应接受严格数据治理与伦理评估,提供充分的解释与干预机制。

五、深化发展

(一)释放数据价值:引导算法算力共优化

教育大模型的发展亟需重视数据的全要素价值。数据本身具有揭示学习过程和行为模式特征,能在模型理解语境和生成响应等方面发挥作用。优质数据可直接提升模型的表现力和泛化能力,支持个性化学习路径优化、教学场景适配等。未来的数据治理不应仅停留在保障数据质量的被动筛选和管理阶段,而应转向更具前瞻性的机制设计,主动地激励社会各界持续生产高质量数据(胡泳等, 2024)。只有构建鼓励多方积极参与、持续供给优质数据的社会环境,才能为教育大模型的发展提供长期、可持续的高质量数据基础。

教育大模型的发展离不开算力、算法与数据的协同优化(李白杨等, 2023)。高效的数据管理与利用策略,可以减轻计算负担,提升训练效率。同时,算法优化也是重要一环,设计高效的算法可以提升教育大模型的应用效果。“数据—算法—算力”的闭环协同能推动教育大模型高质量发展,为教育科研的系统性创新提供坚实支撑。

(二)规范生成数据:平衡生成效率与质量

大模型生成的教学课件、练习题目等可以丰富教学资源库,扩充模型预训练和测评的数据,但可能存在事实错误或潜在偏见(褚乐阳等, 2024)。严格的验证机制可以确保生成内容的准确性和权威性,使其符合教育部门的相关规定和标准。这包括:从事实性、知识性等维度构建严格的评估标准,并建立完善的校验机制确保数据的准确性;对于符合规则且经过验证的优质数据,可通过共享平台开放使用,促进数据的有效流通,丰富模型的认知边界;数据共享平台及时收集各类意见,持续优化数

据质量,提升内容的实用性和适用性。

(三)强化知识支撑:构建图谱驱动新生态

数据治理不仅要关注数据本身的质量,还需加强数据间的逻辑关联。知识图谱通过实体、关系和属性的显式连接,为数据赋予了丰富的语义内涵和逻辑关联,可以降低大模型的“幻觉”风险(Pan et al., 2024)。知识图谱可以通过构建数据的逻辑框架提升数据的结构化程度和可理解性,通过跨领域知识的关联整合促进数据的深度挖掘与知识发现,为大模型提供系统和完整的知识基础。

在数据治理层面,数据管理方应建立数据标注规范,统一知识图谱实体关系的定义和属性描述,确保新增数据能够准确映射到知识图谱框架中;建立知识图谱的质量评估机制,验证知识关联的准确性和完整性。在数据共享层面,建立统一的知识图谱共享平台,有助于促进各教育机构间知识结构的互联互通,实现知识的协同补充和更新;分级授权机制能在保护敏感信息的前提下,最大化知识图谱的共享价值。

(四)构建协同机制:打造数据治理新格局

教育数据的治理与共享离不开多方协同合作,责任清晰、权责对等多元协同治理机制能推动形成“政府引导、机构主责、平台协作、社会监督”的治理格局。政府应发挥引导作用,制定相关政策法规,推动教育数据治理与共享的标准化、法制化建设。教育机构应承担主体责任,提供高质量教育数据,同时增强数据安全意识,规范数据使用行为。平台企业应开发符合教育场景的人工智能产品,为教育数据的安全治理与共享提供技术支撑。科研机构可通过研究评估、方法创新为标准完善和治理优化提供理论依据与实践反馈。媒体、公众和第三方评估机构可以参与教育数据使用的合规性监督,提升数据治理的透明度与公众信任。

构建多方协同治理与激励引导并重的机制,是打造教育数据治理与共享新生态、释放数据社会价值与教育潜能的必由之路。提升协作效能的关键在于通过成立教育数据共享联盟,汇聚多方资源与力量,共同推进数据共享的规范化和规模化。跨领域的数据协作,则需要鼓励教育数据与就业、医疗、社会保障等领域的数据融合共享,推动教育治理与社会服务的联动,实现更广泛的价值共创。此

外,设计合理的共享激励机制和价值反馈机制,也是调动教育机构和学校参与数据共享积极性的有效手段。

[参考文献]

- [1] 包冬梅,范颖捷,李鸣(2015). 高校图书馆数据治理及其框架[J]. 图书情报工作, 59(18): 134-141.
- [2] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & Agarwal, S.(2020). Language models are few-shot learners[J]. Advances in neural information processing systems, 33(1): 1877-1901.
- [3] Centers for Medicare & Medicaid Services(2022). A quick guide to data catalog[EB/OL]. [2025-06-03]. <https://www.cms.gov/files/document/data-catalog-quick-reference-guide.pdf>.
- [4] 陈桂香,吴晨璐(2023). 我国高校数据治理体系要素构成、存在问题及解决对策——活动理论视角[J]. 高校教育管理, 17(3): 63-75.
- [5] 陈丽,徐亚倩,杨阳(2024). “互联网+教育”的方法论:共享驱动创新[J]. 中国远程教育, 44(1): 25-34.
- [6] 程娅(2022). 面向未来:欧盟数据治理框架的要素分析与经验启示[J]. 数字图书馆论坛, 18(12): 47-53.
- [7] Cheong, L. K., & Chang, V. (2007). The need for data governance: A case study[C]// Proceedings of the 18th Australasian Conference on Information Systems Doctoral Consortium (ACIS 2007). Toowoomba, Australia. IEEE Computer Society: 100.
- [8] Chowdry, H., Crawford, C., Dearden, L., Goodman, A., & Vignoles, A.(2013). Widening participation in higher education: Analysis using linked administrative data[J]. Journal of the Royal Statistical Society Series A: Statistics in Society, 176(2): 431-457.
- [9] 褚乐阳,潘香霖,陈向东(2024). AI大模型在教育应用中的伦理风险与应对[J]. 苏州大学学报(教育科学版), 12(1): 87-96.
- [10] Council of the European Union(2024). Artificial Intelligence (AI) Act: Council gives final green light to the first worldwide rules on AI[EB/OL]. [2025-06-03]. <https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>.
- [11] DAMA International (2017). DAMA-DMBOK: Data management body of knowledge[M]. Sedona, AZ: Technics Publications: 30.
- [12] Data Governance Institute(2019). DGI data governance framework[EB/OL]. [2025-06-03]. <https://datagovernance.com/the-dgi-data-governance-framework/>.
- [13] 丁志刚(2014). 论国家治理体系及其现代化[J]. 学习与探索, 233(11): 52-57.
- [14] 董伟玮,秦音(2025). 基于“技术—制度”框架的超大城市政务数据共享体系建构与优化策略[J]. 现代情报, 45(6): 140-152.
- [15] 董晓辉(2021). 活动理论视角下高校教育数据治理体系构成要素研究[J]. 中国电化教育, (3): 79-87.
- [16] EDFacts Data Governance Board(2018). Maximizing ED-Facts data quality: A comprehensive approach[EB/OL]. [2025-06-03]. <https://www.ed.gov/sites/ed/files/about/inits/ed/edfacts/edfacts-data-quality-process-overview.pdf>.
- [17] European Commission(2025). A European strategy for data [EB/OL]. [2025-06-03]. <https://digital-strategy.ec.europa.eu/en/policies/strategy-data>.
- [18] 国家标准委(2017). GBT+30850.4-2017, 电子政务标准化指南第4部分:信息共享[S]. 中华人民共和国国家质量监督检验检疫总局、中国国家标准化管理委员会. <http://c.gb688.cn/bzgk/gb/showGb?type=online&hcno=5C310629704FBBF3F58CB6B3786B25F8>.
- [19] 国家标准委(2018a). GB/T 34960.5-2018, 信息技术服务治理第5部分:数据治理规范[S]. 国家市场监督管理总局, 中国国家标准化管理委员会. <http://c.gb688.cn/bzgk/gb/showGb?type=download&hcno=F3B2108863A2292F5AF0FA645CEE047F>.
- [20] 国家标准委(2018b). GB/T 36073-2018, 数据管理能力成熟度评估模型[S]. 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. <https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=B282A7BD34CAA6E2D742E0CAB7587DBC>.
- [21] 国家标准委(2020). GBT+38664-2020, 信息技术 大数据 政务数据开放共享[S]. 国家市场监督管理总局, 中国国家标准化管理委员会. <https://std.samr.gov.cn/gb/search/gbDetailed?id=A47A713B767914ABE05397BE0A0ABB25>.
- [22] 国家标准委(2024a). GB/T 43697-2024, 数据安全数据分类分级规则[S]. 北京: 中国标准出版社: 2-3. <https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=F0C385EDC38CBF277AEC021F23126ADE>.
- [23] 国家标准委(2024b). GB/T 44109-2024, 信息技术 大数据 数据治理实施指南[S]. 国家市场监督管理总局, 中国国家标准化管理委员会. <https://std.samr.gov.cn/gb/search/gbDetailed?id=19B8508AEDE64286E06397BE0A0A616D>.
- [24] 国家发改委, 国家数据局, 中央网信办, 工业和信息化部, 财政部, 国家标准委(2024). 国家发展改革委等部门关于印发《国家数据标准体系建设指南》的通知[EB/OL]. [2025-01-06]. https://www.gov.cn/zhengce/zhengceku/202410/content_6978809.htm.
- [25] 国家矿山安全监察局(2023). 《智能化矿山数据融合共享规范》[EB/OL]. [2025-01-06]. https://www.chinamine-safety.gov.cn/xw/zl/2018zt/mkjqrjfy/gzdt_01/202307/t20230704_455286.shtml.
- [26] 哈尔滨市大数据中心(2024). DB2301/T 184-2024, 公共数据数据治理规范[S]. 哈尔滨市市场监督管理局. <https://std.samr.gov.cn/db/search/stdDBDetailed?id=2A7AEBB3E3B48461E06397BE0A0AB9FE>.
- [27] 胡新瑞(2025). 教育数字化转型下教育数据的分类治理[J]. 比较教育学报, (1): 3-13.
- [28] 胡泳,刘纯懿(2024). 大语言模型“数据为王”:训练数据的价值、迷思与数字传播的未来挑战[J]. 西北师大学报(社会科学版), 61(3): 43-54.
- [29] 华为技术有限公司, 中国信通院云计算与大数据研究所(2023). 关于公共数据授权运营的数据流通建设白皮书[EB/OL]. [2025-06-03]. <https://13115299.s21i.faiusr.com/61/1/ABUIABA9GAAg7fG1qAYov9Tz-wc.pdf>.
- [30] International Organization for Standardization(2015). ISO/IEC 38500: 2015, Information Technology: Governance of IT for the Organization[S]. ISO/IEC. <https://www.iso.org/obp/ui/#iso:std:iso-iec:38500:ed-2:v1:en>.

- [31] International Organization for Standardization(2017). ISO/IEC 38505-1: 2017, Information technology—Governance of IT—Governance of data—Part 1: Application of ISO/IEC 38500 to the governance of data[S]. ISO/IEC. <https://www.iso.org/obp/ui/#iso:std:iso-iec:38505:-1:ed-1:v1:en>.
- [32] International Organization for Standardization(2022). ISO/IEC 27001: 2022, Information security, cybersecurity and privacy protection—Information security management systems—Requirements[S]. ISO/IEC. <https://www.iso.org/standard/27001>.
- [33] International Organization for Standardization(2024). ISO/IEC 5259-4: 2024, Artificial intelligence - Data quality for analytics and machine learning (ML) -Part 4: Data quality process framework[S]. ISO/IEC. <https://www.iso.org/obp/ui/#iso:std:iso-iec:5259:-4:ed-1:v1:en>.
- [34] Kramcsák, P. T.(2023). Can legitimate interest be an appropriate lawful basis for processing artificial intelligence training datasets?[J]. *Computer Law & Security Review*, 48(1): 105765.
- [35] 李白杨,白云,詹希旎,李纲(2023). 人工智能生成内容(AIGC)的技术特征与形态演进[J]. *图书情报知识*, 40(1): 66-74.
- [36] 李青,韩俊红(2018). 数据治理:提升教育数据质量的方法和途径[J]. *中国远程教育*, 38(8): 45-53+80.
- [37] 李森(2023). 风险防范视阈下生成式人工智能数据安全的治理路径——以GPT类模型为例[J]. *西藏民族大学学报(哲学社会科学版)*, 44(6): 139-145.
- [38] 李新,郑奕珂,杨现民(2024). 心理生理学数据的教育应用:科学释义、实践探索与发展趋向[J]. *现代远程教育*, (5): 44-58.
- [39] 刘三女牙,杨宗凯,李卿(2017). 教育数据伦理:大数据时代教育的新挑战[J]. *教育研究*, 38(4): 15-20.
- [40] 刘洋溪(2024). 迈向敏捷治理:教育数据治理的风险识别与优化路径[J]. *华中师范大学学报(人文社会科学版)*, 63(6): 154-164.
- [41] 罗威,罗准辰,雷帅,程齐凯,陆伟,张瑾,韩涛,冯岩松,韩先培,冯冲,张均胜,刘志辉,乔林波,李东升,许儒红,陈敬一(2020). 智能科学家——科技信息创新引领的下一代科研范式[J]. *情报理论与实践*, 43(1): 1-5+17.
- [42] 马治国,张楠(2024). 通用人工智能的数据风险及法治应对路径[J]. *北京工业大学学报(社会科学版)*, 24(5): 131-142.
- [43] Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H.(2009). Overview and framework for data and information quality research[J]. *Journal of data and information quality (JDIQ)*, 1(1): 1-22.
- [44] 梅傲,谢冰姿(2024). 总体国家安全观视域下我国数据安全主体责任研究[J]. *科技与法律(中英文)*, (4): 30-39.
- [45] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X.(2024). Unifying large language models and knowledge graphs: A roadmap[J]. *IEEE Transactions on Knowledge and Data Engineering*, 36(7): 3580-3599.
- [46] 彭振宇,王羽萱(2025). 突破价值释放瓶颈:多源异构教育数据通用语义模型的构建与应用[J]. *中国电化教育*, (4): 40-47.
- [47] 钱冬明,管珏琪,祝智庭(2013). 数字教育资源共建共享的系统分析框架研究[J]. *电化教育研究*, 34(7): 53-58+70.
- [48] 全国信息技术标准化技术委员会(2017). GB/T 35295-2017, 信息技术 大数据 术语[S]. 中华人民共和国国家质量监督检验检疫总局;中国国家标准化管理委员会. <https://std.samr.gov.cn/gb/search/gbDetailed?id=71F772D823FCD3A7E05397BE0A0AB82A>.
- [49] Regulation, P.(2018). General data protection regulation[J]. *In-touch*, 25(1): 1-5.
- [50] Schildkamp, K., Lai, M. K., Earl, L. (2012). Data-based decision making in education: Challenges and opportunities [M]. Dordrecht, Netherlands: Springer Dordrecht: 23-47.
- [51] 山东省大数据局(2023). DB37/T 4646—2023, 公共数据 数据治理规范[S]. 山东省市场监督管理局. <https://std.samr.gov.cn/db/search/stdDBDetailed?id=03B9A83C2CCE5065E06397BE0A0A4EC9>.
- [52] 山东省大数据局(2024). DB37/T 4755.2—2024, 公共数据共享 第2部分:服务规范[S]. 山东省市场监督管理局. <https://std.samr.gov.cn/db/search/stdDBDetailed?id=21E6D7791CC28328E06397BE0A0AE3FC>.
- [53] 山东省教育信息中心(2019). DB37/T 3600.3-2019, 基础教育管理服务平台 第3部分:数据共享[S]. 山东省市场监督管理局. <https://std.samr.gov.cn/db/search/stdDBDetailed?id=95F08A5658AF45C1E05397BE0A0A3A02>.
- [54] 上海市大数据中心(2020). DB31/T 1240-2020, 公共数据共享交换工作规范[S]. 上海市市场监督管理局. <https://std.samr.gov.cn/db/search/stdDBDetailed?id=B9B07703A19FB186E05397BE0A0ACF77>.
- [55] 苏福根,杨伟平(2025). 质量视角:高等教育数据治理的基本逻辑与实践路径[J]. *教育学术月刊*, (2): 34-40.
- [56] 孙洪涛,郑勤华(2016). 教育大数据的核心技术、应用现状与发展趋势[J]. *远程教育杂志*, 34(5): 41-49.
- [57] 孙旭东,檀昌稳,杨洋,莫蕾钰(2024). 共建共享视角下数智校园的理论框架与数据治理实现路径[J]. *现代教育技术*, 34(8): 132-141.
- [58] 唐颖,刘钰,谢涛(2025). 教育数据隐私保护的困境与突破——基于世界一流高校政策文本的多维分析[J]. *中国远程教育*, 45(4): 69-84.
- [59] 田贤鹏(2020). 隐私保护与开放共享:人工智能时代的教育数据治理变革[J]. *电化教育研究*, 41(5): 33-38.
- [60] 王娟,王书瑶,杨现民,娄方园(2022a). 教育政务数据开放共享的保障框架与路径设计[J]. *电化教育研究*, 43(4): 63-69.
- [61] 王娟,杨现民,高振,王书瑶(2022b). 大数据时代教育政务数据开放共享的监管机制[J]. *现代远程教育研究*, 34(3): 67-75.
- [62] 王利明(2019). 数据共享与个人信息保护[J]. *现代法学*, 41(1): 45-57.
- [63] Wells, D. (2018). The Ultimate Guide to Data Catalogs: Key Things to Consider When Selecting a Data Catalog [EB/OL]. [2025-06-03]. https://s3.amazonaws.com/eckerson/content_assets/assets/000/000/177/original/Data_Catalogs_-_April_2018.pdf?1525198138.
- [64] 吴永和,颜欢,马晓玲(2024). 教育通用人工智能大模型标准体系框架研制[J]. *现代教育技术*, 34(4): 28-36.
- [65] 谢娟(2020). 教育数据治理的伦理框架:价值、向度与路径[J]. *现代远程教育研究*, 32(5): 15-24.
- [66] 徐峰,吴旻瑜,徐萱,任友群(2018). 教育数据治理:问题、思考与对策[J]. *开放教育研究*, 24(2): 107-112.
- [67] 徐晓林,明承瀚,陈涛(2018). 数字政府环境下政务服务数据共享研究[J]. *行政论坛*, 25(1): 50-59.

- [68] 颜世健, 喻国明(2024). 智能方法作为“第五范式”: 人工智能时代科研范式的“新物种”[J]. 学术探索, (1): 34-43.
- [69] 杨现民, 唐斯斯, 李冀红(2016). 发展教育大数据: 内涵、价值和挑战[J]. 现代远程教育研究, (1): 50-61.
- [70] 杨现民, 陈世超, 唐斯斯(2017). 大数据时代区域教育数据网络建设及关键问题探讨[J]. 电化教育研究, 38(1): 37-46.
- [71] 杨现民, 周宝, 郭利明, 杜沁仪, 邢蓓蓓(2018). 教育信息化2.0时代教育数据开放的战略价值与实施路径[J]. 现代远程教育研究, (5): 10-21.
- [72] 杨现民, 王娟, 李新(2023). 加强国家智慧教育平台数据治理: 经验洞察与路径优化[J]. 中国电化教育, (9): 69-75.
- [73] 杨现民, 曾佳尧, 赵路瑶, 李新(2024). 共同体理论视角下研究生优质教学资源共享机制与实施路径设计[J]. 现代教育技术, 34(10): 75-82.
- [74] 杨伟平, 曾德华, 段婷婷, 赵昱(2025). 数字化转型背景下的教育数据分类分级研究[J]. 现代教育技术, 35(1): 89-97.
- [75] 余鹏, 李艳(2018). 大数据视域下高校数据治理方案研究[J]. 现代教育技术, 28(6): 60-66.
- [76] 张绍华, 潘蓉, 宗宇伟(2016). 大数据治理与服务[M]. 上海: 上海科学技术出版社: 143-164.
- [77] 张绍丽, 郑晓齐, 张辉(2017). 基于资源共享的教育大数据信息平台构建及机制研究[J]. 现代情报, 37(12): 90-95+101.
- [78] 张欣(2023). 生成式人工智能的数据风险与治理路径[J]. 法律科学(西北政法大学学报), 41(5): 42-54.
- [79] 张越, 郭玥, 余江(2025). 通用人工智能驱动的科研新范式: 理论与实践[J]. 科学学研究, 43(4): 673-682.
- [80] 赵润娣(2021). 我国政府数据开放分类分级研究——基于开放政府数据平台教育类数据的调查[J]. 现代情报, 41(4): 90-100.
- [81] 中国电子技术标准化研究院(2024). 数据要素流通标准化白皮书[EB/OL]. [2025-01-06]. <https://www.cesi.cn/202405/10096.html>.
- [82] 中国通信标准化协会(2021). 数据治理标准化白皮书[EB/OL]. [2025-06-03]. <https://13115299.s21i.faiusr.com/61/1/ABUIA BA9GAAG7YGHjgYo4PjDtgQ.pdf>.
- [83] 周代数, 魏杉汀(2024). 人工智能驱动的科学第五范式: 演进、机制与影响[J]. 中国科技论坛, (12): 97-107.
- [84] 周东岱, 李振, 刘志勇(2019). 基于流式处理技术的异构教育数据交换方法研究[J]. 现代教育技术, 29(8): 106-112.
- [85] 周力虹, 胡江枫(2025). “进攻与防守”: 基于领地理论的跨组织数据共享障碍研究[J]. 图书情报工作, 69(7): 3-15.

(编辑: 魏志慧)

Research on the Construction of Data Governance and Sharing Standard Framework of Large Models in Education

WU Yonghe¹, CHEN Yuanyuan¹, WU Huina¹ & MA Xiaoling²

(1. Department of Educational Information Technology, East China Normal University, Shanghai 200062, China; 2. Department of Information Management, East China Normal University, Shanghai 200062, China)

Abstract: Under the intelligent paradigm, educational data exhibits characteristics such as heterogeneous sources, unclear authorization, context dependence, and high sensitivity, all of which pose multiple challenges, including collaborative governance, privacy protection, and cross-domain sharing. Focusing on high-quality educational data, this study proposes the connotation characteristics of a three-tier structure—“security–quality–context”—and clarifies the core goals and value orientation for standardized management and efficient utilization of educational data. The study establishes a standard framework for data governance and sharing of educational large models with future development envisioned in four dimensions: optimizing algorithmic computing power, standardizing generated data, strengthening knowledge support, and establishing collaborative mechanisms. The proposed framework is expected to provide robust data support for the pre-training and evaluation of educational AI large models, facilitate their safe, efficient, and trustworthy application, and advance innovation and high-quality development in intelligent education.

Key words: large model in education; data governance; data sharing; high quality education data; standard framework