

面向健康医疗大数据应用的批判性数据素养 内涵研究

宋月丽, 吴莉莉, 陈贵梅, 王筱婷, 柴静*
安徽医科大学 卫生管理学院, 安徽 合肥 231200
DOI:10.61369/EST.2026070042

摘 要 : 健康医疗大数据的快速发展为疾病预测、个性化治疗、药物研发等医学应用场景带来新的可能, 但由此引发诸如算法歧视、隐私泄露、伦理失范等系统性风险。传统数据素养侧重于数据分析与利用的技术能力, 难以应对数据化背景下日益凸显的权力不对称与伦理困境等问题。为此, 本文聚焦批判性数据素养, 研究其在健康医疗应用场景下的内涵结构, 构建了涵盖“数据情境意识、算法批判能力、数据解释能力、伦理法律意识与行动参与能力”的三维框架。研究表明, 健康医疗领域的批判性数据素养不仅关乎数据技能, 更是一种涉及权力识别、伦理判断及公共参与的复合素养。本文所提出的批判性数据素养内涵框架可为医学院校的课程设计、医疗机构的伦理培训及公共卫生领域的数治理提供理论参考。

关 键 词 : 健康医疗大数据; 批判性数据素养; 数据伦理; 算法偏见; 数据治理

Research on the Connotation of Critical Data Literacy for Health Care Big Data Applications

Song Yueli, Wu Lili, Chen Guimei, Wang Xiaoting, Chai Jing*
School of Health Management, Anhui Medical University, Hefei, Anhui 231200

Abstract : The rapid development of health care big data has brought new possibilities for medical application scenarios such as disease prediction, personalized treatment, and drug research and development, but it has also triggered systemic risks such as algorithm discrimination, privacy disclosure, and ethical anomie. Traditional data literacy focuses on the technical ability of data analysis and utilization, which is difficult to deal with the increasingly prominent problems of power asymmetry and ethical dilemma in the context of data. Therefore, this paper focuses on critical data literacy, studies its connotation structure in the health care application scenario, and constructs a five-dimensional framework covering data context awareness, algorithm criticism ability, data interpretation ability, ethical and legal awareness and action participation ability. Research shows that critical data literacy in the field of health care is not only related to data skills, but also a composite literacy involving power identification, ethical judgment and public participation. The connotation framework of critical data literacy proposed in this paper can provide a theoretical reference for the curriculum design of medical colleges, the ethical training of medical institutions and the data governance in the field of public health.

Keywords : health care big data; critical data literacy; data ethics; algorithm bias; data governance

引言

大数据、人工智能等信息技术的快速发展和广泛应用, 给健康医疗领域带来深刻变革, 但也使得该领域的伦理与安全问题日益凸显, 比如数据泄露、未经授权访问的风险问题等。近年来医疗数据泄露事件的数量和频率大幅增加, 由此也引发了广泛的伦理担忧^[1]。另外, 算法偏见导致医疗资源分配不均或种族偏见, 以及当前的数据伦理规范难以跟上技术迭代的速度等问题同样不容忽视^[2]。

面对健康医疗大数据带来的伦理风险与挑战, 传统的“数据素养”(Data Literacy)表现出明显的局限性。因为传统的数据素养主要强调数据采集、清理、分析等全流程中的的相关技能^[3], 往往忽略了数据背后的政治经济属性。在健康医疗领域这种局限性更为突

基金项目: 2025年安徽省教育厅人文社科重点项目(项目编号: 2025AHGXSK30233); 2025年度安徽省省级质量工程项目(项目编号: 2025dzxkc021, 2025aijy105, 2025sx027); 2024年度安徽省省级质量工程项目(2024xqhz021)。

作者简介: 宋月丽(1987-), 女, 讲师, 研究方向为数据素养教育、健康医疗大数据应用

通讯作者: 柴静(1982-), 女, 副教授, 研究方向为数智医疗、数据素养, 邮箱: 2008500017@ahmu.edu.cn

出,算法偏见、隐私侵犯等问题不仅破坏了数据的公平性,更持续加剧了数据主体(包括公民、医疗从业者等)与数据处理者(如大型医疗平台、医疗公司、医保机构等)之间的权力失衡。因此,亟需构建一种批判性数据素养(Critical Data Literacy),以使利益相关者能够识别医疗数据实践中的权力不对称,质疑算法决策的公正性,并在伦理困境中做出审慎判断。

批判性数据素养是一种能够促进对数据实践进行批判性审视,能够将数据置于平台、网络基础设施和社会背景中加以理解的复合能力^[4,5]。但在健康医疗这一特殊敏感领域,批判性数据素养的内涵尚缺乏系统性梳理。本研究将梳理数据素养从工具性技能向批判性实践的概念演进脉络,分析健康医疗大数据面临的核心伦理风险,在此基础上构建批判性数据素养在医疗情境下的内涵维度,以期为医学院校的课程设计、医疗机构的伦理培训及公共卫生领域的治理提供理论参考。

一、批判性数据素养的形成及相关研究

数据素养作为一个学术概念,其内涵在近几十年中经历了持续演变。从最初聚焦于对统计信息的解读与评估,到后来扩展为涵盖数据全流程的综合性能力,再到近年来在批判性视角下被重新审视,这一概念的每一次转向都与技术环境的变迁和社会需求的更替密切相关。在健康医疗领域,数据素养的研究同样经历了从技术能力向批判性实践的渐进式拓展。

(一) 数据素养的概念演进与局限

有关数据素养的学术研究可追溯至统计素养与信息素养两个源头。Gal(2002)将统计素养定义为对统计数据进行解读和批判性评估,以及对其中的信息和观点开展讨论或交流的能力^[6]。随着信息技术普及,信息素养的概念逐渐兴起,其核心观点包括如何识别信息需求、如何获取和评估信息、以及如何合乎伦理地使用信息。

进入大数据时代,数据素养的内涵经历了显著扩展,其核心不仅包括审视统计数据,还要能遵循严谨的数据探究流程来思考和处理数据^[7]。在健康医疗领域,数据素养最初表现为医疗工作者对电子健康档案的操作以及对医疗统计数据进行分析等基础技术能力。随着医疗信息化的深入,这一概念逐渐延伸至对数据质量、数据安全与数据伦理等问题的关注。数据素养对于医疗工作者而言,不仅意味着能够使用数据,更意味着能够理解数据的来源、评估数据的可靠性并在临床决策中审慎地运用数据^[8]。但这仍然没有跳出数据的技术性操作范畴。

事实上,传统的数据素养框架往往默认数据是客观、中立且价值无涉的,其关注焦点集中在个体层面的技能培养。这在数据规模较小、应用场景相对单纯的场景下或许是够用的。但当大数据系统被广泛应用于医疗诊断等涉及公民基本权益的领域时,传统数据素养便难以应对数据化社会中比较突出的权力不对称或结构性不公等问题。

(二) 批判性数据素养的兴起与核心主张

批判性数据素养的兴起,与批判教育学及批判性数据研究的学术脉络密切相关,它更强调要理解数据的社会情境性与权力依赖性^[9]。Tygel与Kirsch(2016)提出批判性数据素养的四项基本能力,包括数据阅读(理解数据如何生成及其背景)、数据处理(将数据转化为信息)、数据沟通(选择恰当方式传达信息)以及数据生产(以合乎伦理的方式创建和发布数据)^[10],他们还

强调数据素养不应止步于技术层面的数据操作,而应服务于社会正义与公民赋权。

Sander(2024)提出的“批判性数据化素养”框架^[3],则包含了三个核心目标,包括促进对数据化的系统性理解、鼓励批判性思维、使学习者能够采取不同形式的行动。在方法论层面,Dangol与Dasgupta(2023)等人的研究结果表明,将批判性数据素养与建构主义学习理论相结合,能够有效帮助学生在数据项目的实践中理解数据的起源与局限,并发展出质疑与行动的能力^[4]。这对健康医疗领域的数字素养教育具有重要启示,因为医疗工作者在接触临床数据时,同样需要理解这些数据是如何在特定的诊疗规范、制度安排和价值取向中被生产出来的。

(三) 健康医疗大数据伦理挑战对批判性数据素养的呼唤

健康医疗大数据因其敏感性与复杂性而面临多重伦理挑战,这些挑战恰恰凸显了批判性数据素养的紧迫性与必要性。隐私泄露风险是最为直观的伦理问题。传统的研究伦理框架往往假定研究者与研究对象之间存在直接的互动关系,但大数据研究中的数据处理器与数据主体之间往往并无直接接触,这使得知情同意等传统保护机制面临失效的风险^[2]。在健康医疗领域,这一问题尤为突出,例如患者可能完全不知晓自己的诊疗数据被用于何种研究或商业目的。算法偏见问题同样不容忽视。有研究表明算法决策往往会来自低收入家庭和社区的人群产生不利影响^[11]。在医疗场景中,这意味着某些群体可能因算法偏见而被系统性地排除在先进诊疗资源之外,或者只能获得质量较低的医疗服务。而想要在健康医疗大数据应用场景中保持技术中立原则,就要求医疗相关工作者不仅能够使用算法工具,还需要具备识别和质疑算法偏见的能力。伦理监管滞后构成了第三重挑战。由于制度更新难以跟上技术迭代的速度,使得监管真空与伦理失范现象同时存在。伦理审查委员会在面对大数据研究时普遍存在专业知识不足和审查标准难以统一的问题^[12]。

正是由于上述挑战的存在,批判性数据素养在健康医疗领域的价值就显得格外突出。因此,构建面向健康医疗大数据应用的批判性数据素养内涵框架,进而提升相关主体的批判性数据素养水平,就成为应对该领域众多伦理风险挑战的迫切需求和重要途径。

二、健康医疗领域批判性数据素养的五维内涵阐释

基于对批判性数据素养形成脉络的梳理,以及对健康医疗大

数据伦理挑战的分析，本文提出面向健康医疗领域的批判性数据素养五维内涵框架，五个维度各有侧重又相互关联，共同构成一个从认知到行动、从个体到集体的素养体系。

（一）数据情境意识

数据情境意识是批判性数据素养的认知基础，要求医疗从业者要能够认识到数据的非中立性及其对情境的依赖性。数据伦理的核心要求之一便是批判性地理解技术环境及其与数据中权力层次的交织关系，而这种理解必须建立在对数据产生和流通情境的深度把握之上^[12]。在健康医疗领域，数据情景意识意识具体体现在以下两个层面。

（1）数据产生的情境性。作为医疗工作者，应能够认识到电子健康档案中的数据并非客观中立的“原始记录”，而是夹杂着临床工作规范、制度安排与权力关系的产物。技术本身也并非价值中立，因其在实际运用中所产生的效果，取决于其所嵌入的社会环境和使用方式^[13]。也就是说医疗数据的形成过程实际上嵌入了临床诊疗的制度流程与特定价值倾向，如果忽视这一情境性，而将医疗数据视为纯粹的“客观事实”，便可能忽略数据本身所携带的偏见与局限。

（2）数据流通的系统性。医疗数据在医院、医疗公司、保险公司及第三方平台之间持续流动，但各方对数据的掌握和理解存在显著的信息不对称。例如，数据处理者可以掌握数据流通的全貌，而数据主体（患者乃至一线医疗工作者）往往对数据被用于何种目的以及流向何处等知之甚少^[12]。因此，培养批判性数据素养的第一步，是帮助医疗从业者建立对数据情境的敏感意识，即是要让他们既看到数据“呈现了什么”，也要有意识去追问数据“遮蔽了什么”。

（二）算法批判能力

算法批判能力是批判性数据素养的核心维度，主要是要求医疗从业者能够对医疗 AI 系统进行审慎评估，而非盲目信任算法输出的结果。这一能力重点体现在偏见识别和黑箱质疑两个层面。

（1）偏见识别。医疗 AI 的训练数据集往往来源于特定医疗机构在特定期内的临床记录，这些数据可能天然缺失某些人群的样本，比如缺少罕见病患者、低收入群体、少数民族等的数据记录，使得算法本身存在结构性偏见^[11]。想要防范算法偏见，就要求医疗从业者具备审查训练数据代表性，以及识别固化在数据中的社会不平等的能力。

（2）黑箱质疑。首先需要认识到算法并非客观中立的“计算器”，其设计与训练过程往往嵌入了开发者的价值选择、目标设定与利益倾向。有研究揭示，大型科技公司常以“黑箱”为由掩盖其数据实践中的权力运作，而且与算法相关的信息往往模糊、碎片化且难以理解^[14]。在健康医疗领域，这一问题的严重性在于算法的内部运算逻辑对临床从业者和患者而言高度不透明，而医疗决策的后果又直接关乎患者的生命健康。因此，医疗工作者在算法逻辑、数据来源与决策依据等维度保持审慎的质疑态度就显得至关重要。

（三）数据解释能力

数据解释能力关注的是医疗从业者如何在数据与分析结果之

间建立有意义的关联，并将这种理解有效传递给非专业受众。具体从以下两个方面加以阐释。

（1）可视化批判。数据可视化是医疗大数据呈现的主要方式之一，但图表也并非全是中立的，其设计选择（如色彩编码、坐标轴刻度、数据聚合方式）本身就包含着特定的呈现逻辑与意图。这意味着医疗工作者还需要具备对可视化结果的批判性阅读能力，也就是不仅要能看懂图表“说了什么”，还要能去追问图表是“如何说的”，以及“可能遗漏了什么”。

（2）对数据不确定性的识别与沟通。医疗数据很少是完全确定的，因为样本偏差、测量误差、数据缺失等问题普遍存在于临床数据之中，这些不确定性因素直接影响着数据分析结果的可靠性与适用范围，这也使得基于医疗数据的分析结果往往带有相当程度的不确定性^[1]。医疗工作者需要能够识别这些不确定性，并在临床决策中加以考量，而不是将数据分析的结果就视为是确定无疑的决策依据。

（四）伦理法律意识

伦理法律意识要求医疗工作者要跳出“合规即合格”的浅层认识，并将伦理原则与法律规范深度融入到数据全生命周期的每一个环节。

（1）全周期伦理合规。医疗数据的处理涉及数据采集、存储、分析、共享、发布等多个环节，每一个环节都面临特定的伦理与法律要求。传统的伦理审查框架往往把主要关注点放在研究启动前对方案进行审核，而忽视了对后续环节的监管。但大数据研究是动态的、不可预测的，这使得上述“一次性”审查的框架难以覆盖数据全生命周期的伦理风险^[2]。在健康医疗领域，这意味着医疗工作者不仅需要在数据收集前要验证其来源的合法性，还要在数据分析过程中持续审查是否可能存在算法偏见，以及在数据发布前确保数据充分标识化，从而将伦理考量嵌入数据处理的每一个环节。

（2）对数据使用边界的判断。数据的使用边界，包括什么数据可以被使用、用于什么目的、在什么条件下使用等，是数据伦理实践中的核心问题^[11]。而健康医疗领域中数据的二次使用同样较为常见，比如将临床诊疗数据用于研究或商业目的等，但这一做法往往超出患者最初同意数据被使用的范围。因此，医疗工作者需要具备判断数据使用是否超出合理边界的能力，避免因数据使用不当而损害患者的权益与信任。

（五）行动参与能力

批判性数据素养仅仅具备上述认知和判断仍然是不够的，还需要将批判性理解转化为具体的参与实践，这一维度强调从个体反思走向集体行动，主要内容包括以下两个层面。

（1）开放数据赋能。开放数据为公众参与数据治理提供了具体载体，但在数据被少数主体垄断的情况下，公民往往只能被动地提供数据，却对数据的流通路径、使用目的及潜在风险等问题知之甚少^[15]。健康医疗领域的开放数据，如医疗资源分布数据、公共卫生监测数据等，可以为公众提供知情的前提。通过接触和分析这些数据，公众有望识别数据实践中可能存在的不公平现象，从而打破信息壁垒，真正参与到数据价值的创造与监督

中来。

（2）参与性数据治理实践。批判性数据素养还应当激发医疗工作者参与数据治理制度建设的意愿与能力。Tygel 与 Kirsch（2016）在阐述批判性数据素养时特别强调“数据生产”这一能力，即个人或群体能够主动创造和发布数据，而非仅作为数据的被动提供者^[10]。在健康医疗语境下，这意味着医疗工作者可以参与伦理委员会的审查工作，参与算法审计的实践，参与医疗数据标准的制定，或通过专业组织对数据政策提出建议。这种参与性的实践将批判性认知转化为制度层面的影响力。

三、结论与展望

本文在梳理数据素养概念演进脉络和健康医疗大数据伦理挑

战的基础上，整合批判性数据研究的相关理论资源，构建了面向健康医疗大数据应用的批判性数据素养五维框架，涵盖数据情境意识、算法批判能力、数据解释能力、伦理法律意识与行动参与能力。五个维度从认知到行动逐层递进，共同构成一个相对完整的素养体系。与传统数据素养侧重于数据操作技术不同，批判性数据素养的核心关切在于揭示数据实践中的权力结构与价值预设。本文所提出的五维框架主要是基于文献整合与理论推导得出，尚未通过实证研究加以检验，另外框架的普适性也有待进一步讨论。后续也将基于该框架进一步开发评估工具，通过质性研究方法深入考察不同岗位从业者在实际工作中面临的数据伦理困境，并探索将批判性数据素养融入医学教育的具体教学策略。

参考文献

- [1]Ahmed, M. M., Okesanya, O. J., Oweidat, M., Othman, Z. K., Musa, S. S., & Lucero-Prisno III, D. E. (2025). The ethics of data mining in healthcare: challenges, frameworks, and future directions. *BioData Mining*, 18(47), 1–16.
- [2]Ferretti A, Ienca M, Sheehan M, et al (2021). Ethics review of big data research: What should stay and what should be reformed?[J]. *BMC Medical Ethics*, 22(1):51.
- [3]Sander, I. (2024). Critical datafication literacy – a framework for educating about datafication. *Information and Learning Sciences*, 125(3/4), 270–292.
- [4]Dangol A, Dasgupta S. (2023). Constructionist approaches to critical data literacy: a review[C]//Proceedings of the 22nd Annual ACM Interaction Design and Children Conference (IDC '23). Chicago: ACM: 112–123.
- [5]Pangrazio L, Selwyn N. (2019). ‘Personal data literacies’ : a critical literacies approach to enhancing understandings of personal digital data[J]. *New Media & Society*, 21(2): 419–437.
- [6]Gal, I. (2002). Meanings , Literacy : Components , Responsibilities Statistical [J]. *International Statistical Review*, 70(1), 1 – 25.
- [7]Louie, J. (2022). Critical data literacy: Creating a more just world with data. National Academy of Sciences Workshop on Foundations of Data Science for Students in Grades K – 12 (September 13 – 14, 2022). Education Development Center.
- [8]L ü dorf V, Mainz A, Meister S, et al. Learning objectives matrix in DIM.RUHR: a didactic concept for the interprofessional teaching of data literacy in outpatient health care[J]. *Healthcare*, 2025, 13(6): 662.
- [9]Pangrazio, L., & Selwyn, N. (2018). ‘Personal data literacies’ : A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, 24(1), 1–19.
- [10]Tygel, A. F., & Kirsch, R. (2016). Contributions of Paulo Freire to a critical data literacy: a popular education approach. *The Journal of Community Informatics*, 12(3), 108–121.
- [11]Atenas J, Havemann L, Kuhn C, et al. Critical data literacy in higher education: teaching and research for data ethics and justice[M]//Raffaghelli J E, Sangr ò A. *Data Cultures in Higher Education: Emergent Practices and the Challenge Ahead*. Cham: Springer, 2023b: 293–311.
- [12]Atenas J, Havemann L, Timmermann C. Reframing data ethics in research methods education: a pathway to critical data literacy[J]. *International Journal of Educational Technology in Higher Education*, 2023a, 20(1): 11.
- [13]Stahl, B. C. (2022). From computer ethics and the ethics of AI towards an ethics of digital ecosystems. *AI and Ethics*, 2(1), 65 – 77.
- [14]Nguyen, D., & Beijnon, B. (2024). The data subject and the myth of the ‘black box’ : Data communication and critical data literacy as a resistant practice to platform exploitation. *Information, Communication & Society*, 27(2), 333 – 349.
- [15]Sander I. What is critical big data literacy and how can it be implemented?[J]. *Internet Policy Review*, 2020, 9(2):1–22.