

健康医疗大数据资产化实现路径研究*

安雨婷^{1,2} 王哲² 李超峰³ 周毅²

¹ 新疆医科大学公共卫生学院 乌鲁木齐 830017 ² 中山大学中山医学院 广州 510080

³ 中山大学附属肿瘤医院 广州 510080

[摘要] 目的/意义 有效利用健康医疗行业大数据的潜在价值, 为我国健康医疗大数据资产化研究提供参考。方法/过程 梳理相关文献资料, 概述我国健康医疗大数据资产化的研究现状, 基于数据生命周期理论, 提出健康医疗大数据资产化的实现路径框架。结果/结论 应着力突破健康医疗大数据的技术困境, 优化确权模式, 完善资产化相关政策, 为健康医疗大数据资产的创新应用提供坚实基础, 推动医疗健康产业高质量发展。

[关键词] 健康医疗大数据; 数据资产化; 数据生命周期

[中图分类号] R-058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2025.04.004

Study on the Realization Path of Healthcare Big Data Capitalization

AN Yuting^{1,2}, WANG Zhe², LI Chaofeng³, ZHOU Yi²

¹School of Public Health, Xinjiang Medical University, Urumqi 830017, China; ²Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China; ³Sun Yat-sen University Cancer Center, Guangzhou 510080, China

[Abstract] **Purpose/Significance** To utilize the potential value of big data in healthcare industry effectively, and to provide references for the research of big data capitalization in healthcare industry in China. **Method/Process** The paper reviews relevant literatures and summarizes the current research status of healthcare big data capitalization in China. Based on the data life cycle theory, the realization path framework of healthcare big data capitalization is proposed. **Result/Conclusion** Efforts should be made to break through the technical dilemma of healthcare big data, optimize the authentic right model, and improve the capitalization policies, so as to provide a solid foundation for the innovative application of healthcare big data assets, and promote the high-quality development of healthcare industry.

[Keywords] healthcare big data; data capitalization; data life cycle

1 引言

在推进健康中国战略的进程中, 数据已成为提

升医疗服务质量的核心驱动力。然而, 数据孤岛化与价值挖掘不足限制了其潜能释放, 亟需系统化的资产化路径以支撑行业高质量发展。2016 年《关于促进和规范健康医疗大数据应用发展的指导意

[修回日期] 2024-11-03

[作者简介] 安雨婷, 硕士研究生; 通信作者: 周毅, 教授, 博士生导师。

[基金项目] 国家重点研发计划项目(项目编号: 2022YFC3601600); 广东省自然科学基金项目(项目编号: 2024A1515011989); 广州市科技计划项目(项目编号: 202206010028)。

见》^[1]明确将健康医疗大数据定义为覆盖个人全生命周期的综合数据。2020 年《关于构建更加完善的要素市场化配置体制机制的意见》进一步将数据认定为现代生产不可或缺的基础要素^[2]。作为生产要素的数据蕴含巨大价值,对社会生产经营至关重要^[3]。通过数据要素化,原始碎片化数据经历“原始数据-数据资源-数据产品”的转变,最终形成“数据资产”,成为社会化生产流通的数据要素^[4]。换言之,数据生产要素即数据资产的一种形式^[5]。在此背景下,人工智能和大数据技术推动数据要素增值,形成新质生产力^[6]。健康医疗数据资产化是医疗健康产业新质生产力发展的关键。当前,健康医疗大数据研究聚焦于规模特征和使用价值^[7],相关资产化研究较薄弱。本文总结国内外健康医疗大数据资产化研究现状,结合数据生命周期理论构建实现路径,为我国健康医疗大数据资产化提供借鉴。

2 大数据资产化

2.1 数据资产

数据相关概念演变映射出经济、技术和社会条件变迁。1974 年 Peterson R E^[8]首次围绕债券提出数据资产的概念。20 世纪 80 年代和 90 年代,美国信息高速公路计划推动全球信息化。Fisher T 等^[9]提出“数据即资产”,将焦点回归数据信息本身。2019 年美国《开放、公共、电子和必要的政府数据法案》正式实施,定义数据资产为整合的数据元素或数据集^[10]。2023 年中国通信标准化协会大数据技术标准推进委员会发布《数据资产管理实践白皮书(6.0)》^[11],指出数据资产是企业拥有或控制,能够带来未来经济利益的数据资源。2024 年 2 月财政部发布《关于加强行政事业单位数据资产管理的通知》^[12],定义行政事业单位数据资产为依法履职或提供公共服务过程中持有或控制、预期能产生管理服务潜力或经济利益的数据资源。

2.2 数据资产化

数据资产衍生自数据资源,代表数据资源的“进阶”^[13]。数据资源通常表现为非标准化、非结

构化,质量不一,针对性开发前价值不显著。数据资产则具备高质量、高使用价值、强融合性及可交易性^[14]。1999 年 Gargano M L 等^[15]指出数据挖掘可从数据资源中提炼有价值的信息成为数据资产。尹西明等^[16]认为数据资产化是通过加工、整合和价值挖掘将数据转化为资产,充分释放数据潜在价值的过程,结果形成具有经济价值的数据资产。数据资产化核心在于数据价值提炼与提升^[14]。数字产业化和产业数字化进程中,通过对数据的收集、分析、管理等实现价值,并转化为现金流^[17-18]。目前,因大数据复杂性及各行业需求不同,尚未形成统一方法,仍在探索相关理论与技术路径^[19]。

3 健康医疗大数据资产化

3.1 健康医疗大数据确权

实现数据资产化,需明确数据的权属^[20]。国际医疗数据保护模式包括基于自主性原则的知情同意模式和为维护正当利益的合理处理模式,其核心原则基本相同,即数据的使用应经数据主体同意,保障个人隐私权与信息权益^[21],在公共利益需要的情况下可例外^[22]。2018 年国家卫生健康委员会发布《国家健康医疗大数据标准、安全和服务管理办法(试行)》《医疗卫生机构网络安全管理办法》,明确医疗机构对健康医疗信息的管理义务^[23-24]。若获得国家相关机构认证和授权,医院将获得对数据资源的独家使用权,并可授权他人或相关企业使用^[25]。

3.2 健康医疗大数据资产化关键步骤

选取“健康医疗大数据”“数据资产化”“医疗数据资产”“medical big data”“healthy big data”等为关键词进行文献检索。叶萌等^[26]认为可以通过数据治理产生数据价值,形成医疗数据资产。王雨晨^[25]提出医疗数据需经历汇集、整理、分析、可视化和安全处理等成为数据资产。Visconti R M 等^[27]针对健康医疗大数据多源特性,提出 BRNN-CHO 框架对患者健康状况分类,涵盖数据源、收集、预分析、预处理和分类阶段。Wang Y 等^[28]通过实证分析 109 个大数据实施案例,证实基于大数据分析

的商业价值模型是数据价值实现的有效工具。张卫东等^[29]基于 Data ONE 模型，提出健康医疗大数据价值挖掘框架，涵盖数据采集、整合、价值挖掘、可视化、管理。姜晓萍等^[30]基于大数据价值链理论，将健康医疗大数据价值的创造过程分解成数据采集、处理、汇聚、分析、共享、应用活动。翟运开等^[19]将数据采集、存储、处理、管理、应用及相关环境因素作为影响健康医疗大数据资产化的 6 个关键维度。不同研究描述健康医疗大数据资产化具体过程虽有差异，但核心思想基本一致。通过提炼和总结关键步骤，发现资产化过程是贯穿大数据生命周期的连续活动。目前，国内外研究将数据生命周期视为从产生到消费的完整数据生命框架^[31]。在数据生命周期中，数据处理分为初始加工和后续再利用两阶段，前者涉及数据收集和处理，后者包括深入分析和数据共享等二次应用^[32]。基于上述研究本文提出健康医疗大数据资产化实现路径框架，见图 1。数据最初为无价值状态，经加工链逐步积累挖掘出价值。在数据加工阶段，整合分散数据至数据池，形成数据资源；预处理后，分类分级加密储存；深入挖掘和分析，提取数据资源中有价值的信息，这一过程是数据资产价值创造的基石，并可能带来潜在的经济回报。再利用阶段，依托完善的数据管理体系，确保数据的可追溯性和信息的准确性；探索拓宽数据的应用领域，促进数据融合，催生新的价值点。通过充分的数据价值实现，形成具有经济效益的健康医疗大数据资产。

3.3 健康医疗大数据资产化具体实现

一是数据汇集。健康医疗大数据采集是价值挖掘的基石^[29]。多源异构健康医疗大数据采集方式包括设备数据收集和 Web 数据爬取。前者可通过特定系统接口，利用 Sqoop、Flume 等分布式数据采集集群技术进行收集；后者可运用 Apache Nutch、Scrapy 等工具定向抓取。收集后需确定数据标准，狭义上指元数据标准^[33]，通常根据应用领域、专指度、通用性和互操作级别划分^[34]。二是数据预处理。主要操作有数据清洗、数据集成、数据归约和数据转换等^[35-36]，为下一步分析和深入挖掘奠定基础。三是数据存储。数据结构决定存储方案，健康医疗大数据包括结构化、半结构化、非结构化和文件数据等^[37]。常见存储方式有分布式文件系统、文档存储、列式存储、键值存储、图形数据库和内存存储。分布式文件系统具有高性能、高扩展和高可用性，是大数据领域最基础、核心的功能组件，可用于存储结构化和半结构化数据。MySQL 常用于存储实时性高、响应快的结构化数据，如临床诊疗和健康监测数据。存储后数据可参考《信息安全技术 健康医疗数据安全指南》(GB/T 39725—2020) 等数据分级管理标准规范进行分类^[38]。健康医疗大数据包含可公开信息和患者敏感信息，安全需求各异。因此，需要合理进行数据分级，对关键追溯性信息执行脱敏操作，对隐私数据设置隐私标记。四是数据分析。这是产生数据价值最主要的环节^[39]，通过聚类与分类、关联分析、深度学习等，挖掘数据关联性或进行预测。五是数据可视化。利用图形技术、分层技术、混合技术等，直观展示处理和聚合后的数据^[40]。六是数据管理。数据质量控制指在数据采集、存储、清洗、分析等过程中，采取相应措施保证数据质量以满足分析的要求^[41]。要从成本、数据质量等维度评估并提升数据产品的应用价值^[42]。七是数据应用。可以使健康医疗大数据资产共享发挥更大价值。数据共享方式包括线下共享、基于共享数据库的在线共享、基于请求和反馈的共享^[43]。数据资源经过以上过程形成数据资产后，可进入数据流通交易市场^[44]，并选择场内或场外交易方式。整个框架的具体技术实现，见图 2。

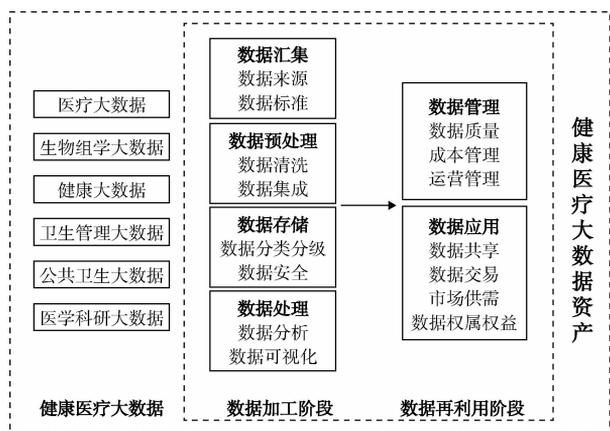


图 1 基于数据生命周期的健康医疗大数据资产化实施路径框架

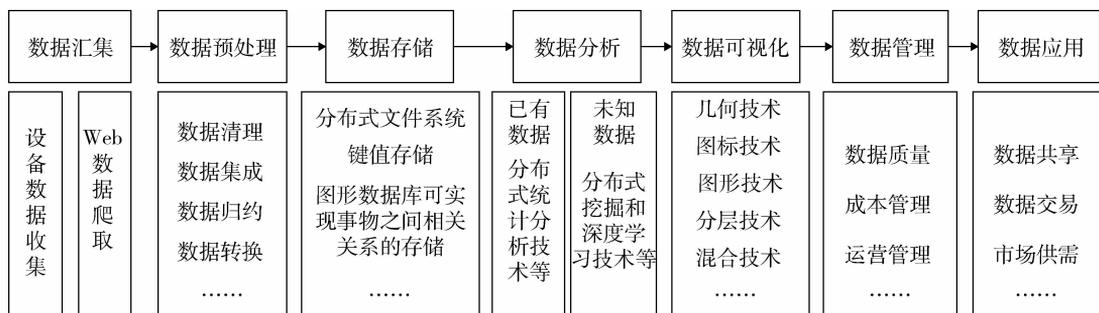


图 2 基于数据生命周期的健康医疗大数据资产化技术流程

4 健康医疗大数据资产化挑战与展望

4.1 技术创新推动健康医疗大数据价值挖掘

健康医疗大数据是典型的多源异构大数据^[37]。不同机构管理信息系统结构不一，导致数据标准、格式差异较大。其中，约 80% 的医疗数据是自由文本构成的非结构化数据，其价值未充分挖掘^[29,44]。尽管已有一些科研和应用基于现有数据产出，但成果较少，且多停留在研究阶段，未进入临床实践^[45]。技术创新将是解决健康医疗大数据技术困境的关键，需提升非结构化数据处理能力，实现数据高效结构化、标准化，挖掘数据价值，为临床实践和科研提供有力支持。

4.2 优化健康医疗大数据确权模式

在健康医疗服务的智能化时代，每个使用者都是数据信息的潜在收集终端^[46]。然而，随着数据流通的频率、深度和广度不断增加，个人信息在流动过程中已悄然超越了其原始生产主体——患者的掌控范围^[47]。健康医疗大数据利用涉及个体权利、企业产权、医疗数据治理及社会公共利益等诸多问题^[23]，传统模式的知情同意原则对数据主体的保护力度略显不足^[48]。未来应构建多方参与的数据确权体系，涵盖政府、医疗机构、企业、患者等，通过法律法规、技术手段和市场机制，实现数据权益的合理分配和有效保护。

4.3 加快医疗数据资产化平台构建

在政策研究层面，我国健康医疗大数据战略聚

焦基础设施建设、应用、服务、保障体系、研究及人才培养 6 大板块^[49]。在实践层面，一方面，多数医院普遍将资源集中于完善自身系统升级，对数据资产化的价值认知与资本投入力度较为有限^[50]；另一方面，医疗机构与企业合作仍局限于科研和临床技术开发的传统模式^[51]，未能充分释放医疗数据的商业潜力。因此建议构建“政府主导 - 医疗机构协同 - 企业参与”的健康医疗大数据共享平台，通过推动医疗数据的商业开发和应用，促进医疗数据在科研、临床、产业等领域的价值转化，实现共赢发展。

5 结语

我国健康医疗大数据资产化实现路径尚未形成一致标准。尽管先进技术的应用为健康医疗大数据资产化提供了坚实的支撑，但数据多源异构特性日益复杂，具体价值实现路径需要持续优化，包括整合多源数据、开发高效精确的数据分析技术、构建可靠的管理系统以及结合不同场景的应用。同时需改进数据确权机制，完善相关政策体系，为我国健康医疗大数据资产的创新应用提供坚实基础，推动医疗健康产业高质量发展。

作者贡献：安雨婷负责论文撰写；王哲负责论文修订；李超峰、周毅负责提供指导。

利益声明：所有作者均声明不存在利益冲突。

参考文献

- 1 关于促进和规范健康医疗大数据应用发展的指导意见 [J]. 中华人民共和国国务院公报, 2016 (19): 24 - 28.
- 2 关于构建更加完善的要素市场化配置体制机制的意见

- [J]. 中国注册会计师, 2020 (5): 6.
- 3 崔云. 数字技术促进新质生产力发展探析 [J]. 世界社会主义研究, 2023, 8 (12): 97-109, 120.
- 4 张帆, 李春光. 数据流通交易平台的全生命周期治理路径研究 [J]. 学习与实践, 2022 (5): 78-84.
- 5 刘雅君, 张雅俊. 数据要素市场培育的制约因素及其突破路径 [J]. 改革, 2023 (9): 21-33.
- 6 加快发展新质生产力 扎实推进高质量发展 [N]. 人民日报, 2024-02-02 (1).
- 7 沈慧煌, 赵静, 马涵彬, 等. 我国健康医疗大数据政策文本分析——基于政策工具视角 [J]. 卫生经济研究, 2021, 38 (8): 11-15, 18.
- 8 PETERSON R E. A cross section study of the demand for money: the United States, 1960—62 [J]. Journal of finance, 1974, 29 (1): 73.
- 9 FISHER T, BOOKS X I. The data asset: how smart companies govern their data for business success [M]. Hoboken: John Wiley & Sons, 2009.
- 10 翟军, 李昊然, 孙小荃, 等. 美国《开放政府数据法案》及实施研究 [J]. 情报理论与实践, 2024, 43 (8): 202-207, 177.
- 11 大数据技术标准推进委员会. 数据资产管理实践白皮书 [EB/OL]. [2024-02-17]. <http://221.179.172.81/images/20230104/12651672818383015.pdf>.
- 12 关于加强行政事业单位数据资产管理的通知 [EB/OL]. [2024-09-20]. https://www.moj.gov.cn/pub/sfbgw/fzgz/fzgzqt/fzgzzbw/202403/t20240318_496236.html.
- 13 王玉林, 高富平. 大数据的财产属性研究 [J]. 图书与情报, 2016 (1): 29-35, 43.
- 14 陆威文, 苟廷佳. 数据要素资产化的理论逻辑与实践进路——基于对数据资产内涵与价值规律的认识 [J]. 企业经济, 2023, 42 (4): 28-39.
- 15 GARGANO M L, RAGGAD B G. Data mining - a powerful information creating tool [J]. OCLC systems & services, 1999, 15 (2): 81-90.
- 16 尹西明, 林镇阳, 陈劲, 等. 数据要素价值化动态过程机制研究 [J]. 科学学研究, 2022, 40 (2): 220-229.
- 17 ARUN K S, GANESH C D. Big data analytics [M]. New York: Chapman and Hall/CRC, 2017.
- 18 张妮. 数据要素价值化: 理论缘起、难点与对策 [J]. 吉首大学学报 (社会科学版), 2024, 45 (5): 52-59.
- 19 翟运开, 刘冰琳, 王宇, 等. 数据生命周期视角下医疗健康大数据资产化影响因素研究 [J]. 情报杂志, 2024, 43 (1): 183-190.
- 20 尤建新, 徐涛. 数字资产化是大势所趋 [J]. 创新世界周刊, 2021 (7): 22-23.
- 21 徐实. 医疗数据权属配置的双重差异化: 多元利益与“三权分置” [J]. 成都理工大学学报 (社会科学版), 2022, 30 (4): 63-75.
- 22 高玉玲, 徐咏军. 公共卫生安全下健康医疗大数据使用中的法律问题及应对 [J]. 中国卫生事业管理, 2021, 38 (12): 918-921.
- 23 国家健康医疗大数据标准、安全和服务管理办法 (试行) [EB/OL]. [2024-10-14]. <http://www.nhc.gov.cn/mohwsbwstjxxzx/s8553/201809/f346909ef17e41499ab766890a34bff7.shtml>.
- 24 医疗卫生机构网络安全管理办法 [EB/OL]. [2024-10-14]. https://www.gov.cn/zhengce/zhengceku/2022-08/30/content_5707404.htm.
- 25 王雨晨. 医院并购中数据资产的评估 [D]. 武汉: 中南财经政法大学, 2023.
- 26 叶萌, 王岩, 朱森. 数据资产视角下医学院校人员信息数据管理的探索与实践——以首都医科大学教务系统为例 [J]. 医学教育管理, 2022, 8 (6): 760-764.
- 27 VISCONTI R M, MOREA D. Healthcare digitalization and pay-for-performance incentives in smart hospital project financing [J]. International journal of environmental research and public health, 2020, 17 (7): 2318.
- 28 WANG Y, HAJLI N. Exploring the path to big data analytics success in healthcare [J]. Journal of business research, 2017, 70: 287-299.
- 29 张卫东, 陈希鹏, 杨斯涵. 健康医疗大数据价值挖掘分析框架构建 [J]. 图书情报工作, 2023, 67 (15): 35-43.
- 30 姜晓萍, 郭宁. 卫生服务数字化转型政策的价值与工具——基于健康医疗大数据的政策分析 [J]. 学海, 2023 (1): 103-113, 146.
- 31 ARASS M E, SOUISSI N. Data lifecycle: from big data to smart-data [C]. Marrakesh: 2018 IEEE 5th International Congress on Information Science and Technology (CIST), 2018.
- 32 吴丹, 马乐. 基于可穿戴设备的医疗健康数据生命周期管理与服务研究 [J]. 信息资源管理学报, 2018, 8 (4): 15-27.
- 33 卢小宾, 洪先锋, 蒋玲. 智慧图书馆数据标准体系研究 [J]. 图书情报知识, 2021, 38 (4): 50-61.
- 34 马琳, 邓宏勇. 国内外临床研究元数据标准: 特点及现状 [J]. 中国循证医学杂志, 2023, 23 (4): 478-484.

- 35 HULSEN T, JAMUAR S, MOODY A, et al. From big data to precision medicine [J]. *Frontiers in medicine*, 2019, 6 (3): 1–14.
- 36 WELCH G, VON RECKLINGHAUSEN F, TAENZER A, et al. Data cleaning in the evaluation of a multi-site intervention project [EB/OL]. [2024-10-14]. <https://up-j-gemgem.ubiquityjournal.website/articles/10.5334/egems.196>.
- 37 陈圣杭, 胡晶, 包耿琿, 等. 健康医疗大数据资源开放对策研究 [J]. *福建电脑*, 2020, 36 (4): 51–54.
- 38 汪火明, 孙润康, 任宇飞, 等. 基于数据分级的医疗大数据中心数据安全策略研究 [J]. *中国医院管理*, 2022, 42 (10): 64–67.
- 39 JAWAD W K, ABBAS M. AI-bakry. Big data analytics: a survey [J]. *Iraqi journal for computers and informatics*, 2023, 49 (1): 41–51.
- 40 GOZZOLI N, SALVATORE F P, FACCILONGO N, et al. How can big data analytics be used for healthcare organization management? Literary framework and future research from a systematic review [J]. *BioMed central health services research*, 2022, 22 (1): 809.
- 41 GLIGORUEVIC V, IAALOD – DOGNIN N, FRZULJ N. Integrative methods for analyzing big data in precision medicine [J]. *Prateomics*, 2016, 16 (5): 741–758.
- 42 黄倩倩, 王建冬, 陈东, 等. 超大规模数据要素市场体系下数据价格生成机制研究 [J]. *电子政务*, 2022 (2): 21–30.
- 43 肖媛, 卢雅雯, 吕智慧, 等. 生物医疗大数据隐私安全保障机制研究 [J]. *计算机应用与软件*, 2021, 38 (2): 318–322.
- 44 郭建. 健康医疗大数据应用中的伦理问题及其治理思考 [J]. *自然辩证法研究*, 2020, 36 (3): 85–90.
- 45 许文韵. 健康医疗大数据中心建设实践与思考 [J]. *医学信息学杂志*, 2020, 41 (8): 48–51, 56.
- 46 刘士国, 熊静文. 健康医疗大数据中隐私利益的群体维度 [J]. *法学论坛*, 2019, 34 (3): 125–135.
- 47 迈尔舍·恩伯格, 肯尼思·库克耶. 大数据时代生活、工作与思维的大变革 [M]. 杭州: 浙江人民出版社, 2013.
- 48 韩旭. 健康医疗数据共享的法律规制研究 [D]. 杭州: 浙江工商大学, 2023.
- 49 李后卿, 印翠群, 樊津妍. 中国健康医疗大数据国家战略发展研究 [J]. *图书馆*, 2019 (11): 30–37.
- 50 顾一纯, 何达, 黄佳好, 等. 基于 PMC 指数模型的我国健康医疗大数据发展的政策环境研究 [J]. *中国卫生政策研究*, 2022, 15 (4): 45–51.
- 51 解媛媛, 张海斌. “业财融合”视角下医院数据资源价值提升的探索 [J]. *中国卫生经济*, 2023, 42 (5): 86–89.

(上接第 14 页)

- 26 BOLLEN K A, STINE R A. Bootstrapping goodness-of-fit measures in structural equation models [J]. *Sociologica methods and research*, 1992, 21 (2): 205–229.
- 27 杜运周, 贾良定. 组态视角与定性比较分析 (QCA): 管理学研究的一条新道路 [J]. *管理世界*, 2017 (6): 155–167.
- 28 甄宓, 刘彦辉, 裴月莹, 等. 信息生态视角下高校大学生学术信息搜寻行为的影响因素与组态路径研究 [J]. *情报科学*, 2024, 42 (10): 155–162.
- 29 金帅岐, 李贺, 沈旺, 等. 用户健康信息搜寻行为的影响因素研究——基于社会认知理论三元交互模型 [J]. *情报科学*, 2020, 38 (6): 53–61, 75.
- 30 周杰, 李小平, 胡德华. 少数民族地区高血压患者健康信息查询行为特征及其影响因素 [J]. *医学信息学杂志*, 2019, 40 (7): 73–76, 85.
- 31 熊皇, 梁亦昆. 与信息偶遇: 媒介化社会中乡村老年群体的在线健康信息获取行为研究 [J]. *现代传播 (中国传媒大学学报)*, 2022, 44 (12): 38–47.
- 32 ZHAO Y C, ZHAO M, SONG S. Online health information seeking among patients with chronic conditions: integrating the health belief model and social support theory [J]. *Journal of medical internet research*, 2022, 24 (11): e42447.
- 33 WANG Z, FAN Y, LV H, et al. The gap between self-rated health information literacy and internet health information-seeking ability for patients with chronic diseases in rural communities: cross-sectional study [J]. *Journal of medical internet research*, 2022, 24 (1): e26308.
- 34 李旭光, 姜富强, 周力虹, 等. 信息生态视角下高校学生网络健康信息搜寻行为影响因素及互动关系研究 [J]. *信息资源管理学报*, 2021, 11 (2): 85–96.