

人工智能驱动的循证健康信息传播系统构建研究

侯震¹，童惟依¹，邓靖飞¹，李扬^{1*} 卣

(中国医学科学院医学信息研究所，北京 100020)

[摘要] 目的/意义 构建基于人工智能的循证医学健康信息传播系统，实现健康热点信息的自动发现与分类，文献证据的分析与智能化数据关联，健康科普信息的智能化生成。
方法/过程 从系统的设计需求出发，从数据层、技术层、应用层、服务层 4 个层面介绍系统的详细架构。
结果/结论 基于人工智能的循证医学健康信息传播系统的构建，打通了热点发现、识别、分类—循证医学证据关联—健康科普内容智能创作的完整研究链条，对提升公众健康信息认知，促进公共卫生水平提高具有实际意义。
[关键词] 人工智能；循证医学；健康信息传播
[中图分类号] R-058 **[文献标识码]** A **[DOI]**

Research on the Construction of an Evidence-Based Health Information Dissemination System Driven by Artificial Intelligence

HOU Zhen¹, TONG Wei-yi¹, DENG Jing-fei¹, LI Yang^{1*}

Chinese Academy of Medical Sciences Institute of Medical Information, Beijing 100020, China

[Abstract] Purpose/Significance To construct an artificial intelligence-based evidence-based medicine (EBM) health information dissemination system that enables the automatic detection and classification of trending health topics, the analysis of literature evidence, and the intelligent association of data for the automated generation of health science popularization content.
Method/Process Starting from the system design requirements, this study introduces the detailed architecture of the system from four levels: the data layer, the technology layer, the application layer, and the service layer.
Result/Conclusion The artificial intelligence-based evidence-based medicine health information dissemination system can effectively realize a complete research chain covering hotspot discovery, recognition, and classification; EBM evidence association; and intelligent creation of health popularization content. This system holds practical significance for improving public understanding of health information and promoting the advancement of public health.
[Key words] artificial intelligence ; evidence-based medicine; health communication

1 引言

随着公众健康意识显著提升，人们对高质量、科学性强且易于理解的健康科普内容的需求日益增长^[1]。然而，当前健康信息传播领域仍存在信息质量参差不齐、传播渠道碎片化以及内容可信度不足等问题^[2]。循证医学作为以科学证据为基础的医学实践模式，为健康科普内容的科学性与规范化提供了坚实支撑^[3]。与此同时，人工智能技术的发展为健康信息的生成、筛选与个性化推荐带来了新的可能，使健康科普的精准传播成为现实^[4]。因此，构建一个人工智能驱动的循证健康信息传播系统，不仅有助于提升公众获取健康信息的效率与质量，也对促进全民健康素养建设、推动健康传播模式的智能化与科学化转型具有重要意义^[5]。

本文提出并构建了一种基于人工智能的循证医学健康信息传播系统，该系统以人工智能大模型为核心工具，以循证医学证据支持为视角，通过健康热点话题的自动发现、研究证据文献的科学遴选以及健康知识的深度融合，实现信息采集到内容生成、传播的智能化闭环，打通了健康信息传播的“数据—知识—内容”通道。该系统的应用推动学术资源高效转化为优质科普内容，提升科技成果的传播效率和社会影响力，对公众健康信息的认知

卣 基金项目：国家重点研发计划“常见多发病防治研究”“十四五”重点专项“脓毒症多器官功能障碍早期识别和动态风险预警体系研究” 项目编号：2021YFC2500800 作者简介：侯震（1991-），男，硕士，助理研究员，研究方向为健康信息传播，数据可视化分析研究等。

*通讯作者：李扬（1978-），女，博士，副研究员，研究方向为情报学、健康传播等。

卣

10.12201/bmr.202610.00043V1

提升具有积极促进作用。

2 相关研究进展

2.1 生成式人工智能技术的发展

生成式人工智能技术是指算法基于大量的预训练数据，根据提示条件自动生产新内容的人工智能技术^[6]。在 20 世纪 80 年代，研究人员尝试以神经网络模型为基础，基于输入的提示内容，生成少量的文本数据，受限于当时的计算机硬件技术及训练数据规模，该时期生成式内容的质量较差。2014 年 Ian Goodfellow^[7]等研究人员提出一种模型训练框架，该框架通过对抗训练，优化样本质量，从而提升生成式图像的质量与分辨率，该模型是人工智能在图像生成领域的重要突破。2017 年随着 Transformer 模型的提出^[8]，生成式人工智能技术迎来了里程碑式的变革。Transformer 模型显著改进了模型训练机制与无监督训练能力，在 GPU 硬件的加持下，其成为生成式文本内容的核心基础。2020 年 OpenAI 公司发布 GPT-3 模型，其基于 Transformer 架构，具备较强的大规模无监督学习能力，可以完成文本自动生成、文章续写、对话等功能。2021 年随着 DALL E、ALIGN 等多模态生成模型的应用^[9]，用户通过提示词及文本引导，即可完成高质量图片、视频内容的生成。

2.2 循证医学在健康传播中的应用

目前国内外已有大量研究学者、机构以循证医学为框架，指导健康传播领域的相关研究及实践工作。在医学科研成果向公众传播方面，由国际非营利组织 Cochrane Collaboration 构建并维护 Cochrane 系统综述平台（Cochrane Systematic Review）^[10]，通过系统地收集、分析关于某个健康领域的高质量研究，得出严谨、可靠的科学结论，用通俗易懂、视觉吸引的方式制作成短视频，让公众更容易理解和应用健康知识。在健康新闻传播领域，李海君等^[11]研究人员发现医学媒体工作者通过系统性学习循证医学框架，其在健康谣言认知和甄别能力方面具有显著提升，从而提高健康信息创作及传播工作效率。在健康教育与行为干预领域，中国医师协会协同北大医学部药物评价中心、中国控烟协会等机构的专家团队，通过循证医学方法制定控烟健康传播指南，提高控烟健康信息传播的有效性^[12]。

2.3 人工智能在健康信息内容生成中的应用

随着公众对健康科普信息的数量、质量、多样性等需求日益增加，传统的健康信息制作模式：依赖专家人工撰写、编辑，多媒体团队加工^[13]，常常面临生产效率低、传播形式单一、内容更新滞后等问题，难以满足公众的需求^[14]。随着人工智能技术的快速发展，其为健康科普内容制作带来了效率、多样性和趣味性的极大提升^[15]。陈娟^[16]等研究人员利用 DeepSeek，针对糖尿病、高血压等 11 种疾病的严重性、患病人群、治疗方案等特征生成健康科普文章，通过文章说服力、公众阅读行为意愿等维度评价生成式健康科普文章的传播效果，研究发现人工智能在撰写患病原因、疾病风险、患病人群特征等方面效果良好，可以协助科普创作者提升工作效率，在诊疗方案、健康行为指导等方面仍需改进。张静^[17]等研究人员基于权威科学的健康科普文案，通过人工智能生成面向多场景、语言风格、目标受众的多样性健康科普口播文案。在保证内容科学完整的前提下，人工智能提升了健康科普创作的多样性。童博^[18]等研究人员在人工智能模型的辅助下，以循证健康证据为基础，生产健康科普作品。经 DISCERN 评审，其生成式健康科普作品的可信性、干预方案内容质量、总体内容质量均较为优秀。

以上研究从多角度验证人工智能模型在健康科普内容创作及传播领域的的能力。但由于人工智能模型训练语料的不确定性，在健康科普内容生成过程中，未经干预的人工智能模型极易生产出幻觉内容误导公众。因此本研究通过训练语料规范筛选、人工智能模型参数调整、生成内容人工审核等多方式，保证生成式健康信息内容的准确性、权威性。

3 系统设计与构建

3.1 设计基础

3.1.1 目标用户

系统由中国医学科学院医学信息研究所开发并运营，其依托权威的生物医学文献资源，建立健康热点话题与研究证据文献之间的动态关联机制。通过人工智能技术，及时发现并响应社会关注的健康热点，科学筛选与之相关的高质量研究证据文献，促进学术资源的科普化转化。系统旨在为健康科普创作人员、科普媒体记者、科研人员、临床工作者等用户提供权威的循证医学支持，协助其开展面向公众的医学知识普及和健康热点问题的循证解读、健康科普信息制作，推动高水平科普资源的开放共享。

3.1.2 设计原则

（1）科学性与权威性：系统以权威生物医学文献为基础，确保健康信息的科学性、准确性和可靠性，强化证据链支撑，杜绝虚假或误导性内容的传播。

（2）动态关联与实时更新：通过自动化、人工智能技术，实现健康热点话题与研究证据文献的动态关联和实时更新，确保信息的时效性和前沿性，及时响应社会健康信息需求。

（3）多元融合与资源共享：整合多渠道、多类型的健康知识及科普资源，促进学术资源与科普资源的深度融合，推动高质量健康信息的开放共享与互惠共赢。

（4）智能化与高效性：采用内容自动识别、内容自动生成等智能技术手段，提高健康信息的筛选、处理、生成的效率，降低人工干预成本，提升系统整体运行效能。

3.1.3 设计思路

系统整体遵循“健康热点发现—证据文献遴选—知识融合与传播”主线，形成完整的信息流转与服务链条。依托社交媒体等多元数据源，构建健康热点话题发现工具，自动采集并识别社会关注度高的健康热点信息。其次，基于文献研究类型、发表时间、期刊影响因子等多维度指标，设计科学合理的研究证据文献遴选策略，实现健康热点与生物医学文献资源的高效动态关联。系统对筛选出的高质量文献进行深度融合，结合健康知识与科普资源，构建权威、系统的健康知识库和科普资源库。系统为科研人员、临床工作者及领域专家提供面向公众的医学概念降维服务、健康科学知识普及服务和健康热点问题循证服务，助力高质量健康信息的传播与社会效益的提升。整体设计注重技术创新与社会价值的有机结合，力求实现健康信息传播的科学化、智能化和高效化，推动学术成果的广泛应用与社会共享。

3.2 系统架构设计

本系统采用分层架构设计，整体分为数据层、技术支撑层、应用层和接口服务层，各层协同配合，保障系统的高效运行和可扩展性。系统支持主流浏览器客户端访问，并严格遵循标准规范体系、安全保障体系和运维管理体系。各层设计如图 1 所示

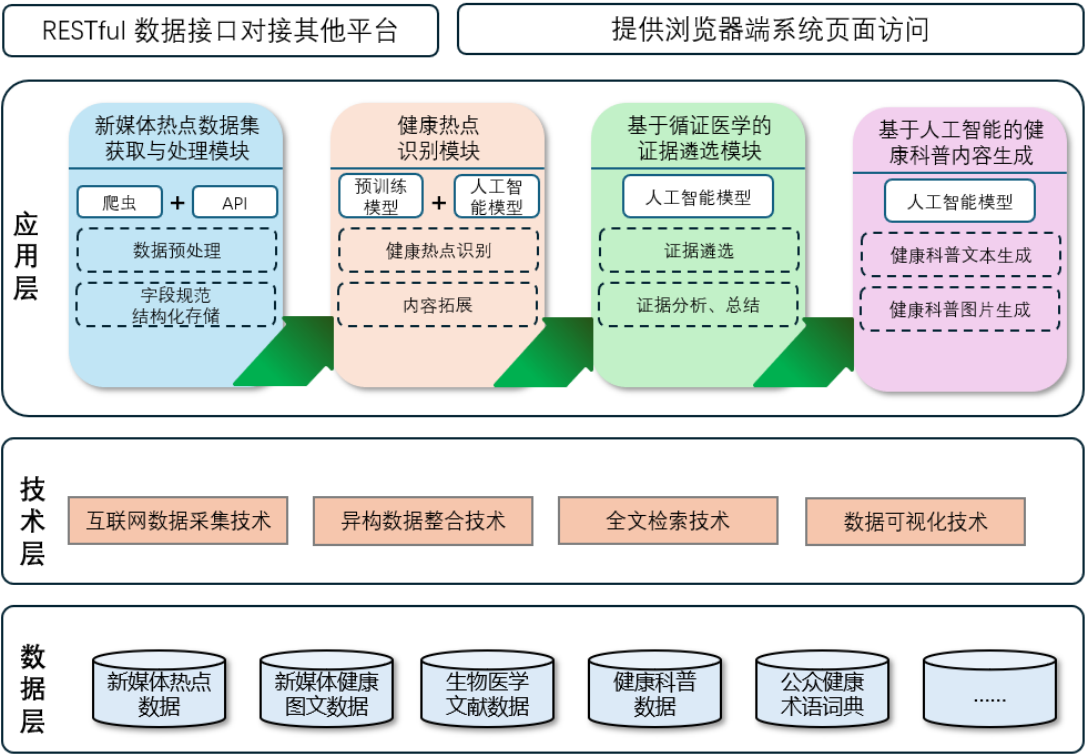


图 1 整体架构图

3.2.1 数据层

数据层负责多源异构数据的采集、存储与管理，对现有元数据以及系统所需的业务数据进行存储，对现有数据信息进行数据清洗、整合、抽取和验证等功能，最后保存到本地数据资源库；对处理好的数据资源进行发布存储至索引库以提供检索服务；将处理好的实例数据存储在数据库中。该层涵盖新媒体热点数据、新媒体健康图文数据、生物医学文献数据、健康科普数据、公众健康术语词典等多类型数据资源。采用 MySQL、Neo4j、Solr、Redis 等多种数据库和索引技术，保障数据的完整性和高效检索。

3.2.2 技术层

技术层为系统开发和运行提供支撑，该层包括互联网数据采集技术、异构数据整合技术、全文检索技术、数据可视化技术等。采用 J2EE、Spring、MyBatis、AJAX 等主流技术框架，实现分布式部署、组件化开发和高效的前后端交互。通过这些技术手段，实现数据的高效整合、检索、分析和可视化展示，支撑知识的深度挖掘与智能应用。

3.2.3 应用层

应用层是系统的核心业务层，负责数据的清洗、分类、存储及知识的发现与展示。主要包括以下功能模块：

新媒体热点数据集获取与处理：实现新媒体热点信息采集和信息分类，提升数据获取的广度和深度。

健康热点识别：通过知识词条筛选、知识发现及信息关联，实现知识的深度挖掘和资源整合。

证据遴选与知识呈现：对健康热点证据进行遴选、分类、分析和概要生成，最终实现知识的可视化展示。

基于人工智能的健康科普内容：基于本地部署的人工智能模型，以健康热点、医学证据、健康知识为数据资源，根据用户提示词信息，自动生成健康科普创作脚本、图文内容。

应用层针对现有的数据资源，利用人工智能手段实现对资源数据的深层次挖掘处理，实现数据的标准化处理和清整。基于 Spring MVC 和 RESTful API 设计，支持多终端访问和

10.12201/bmr.202610.00043V1

业务扩展，提升系统的灵活性和可维护性。

3.2.4 服务层

服务层通过 RESTful API 对外提供统一的数据和知识服务接口，支持客户端（如 IE10+、360、谷歌、火狐等主流浏览器）访问。该层实现了系统各业务模块与外部系统的高效集成与数据交互，集成 Solr 全文检索、Neo4j 图数据库等服务，保障了系统的高可用性和安全性。

3.3.系统模块构建

3.3.1.新媒体热点数据集获取与处理模块

该模块以协作式爬虫协同 API 接口方式，定期抓取微博、微信等自媒体平台的热榜数据。协作式爬虫作为该模块数据集获取的主要方式，无需平台方授权，可根据自定义设置灵活抓取自媒体热榜信息。当出现数据获取不稳定、字段项频繁缺失、抓取模版失效等情况时 API 接口作为辅助方式，保证数据获取的稳定性。通过以上两种方式完成热点数据集采集后，将数据进行统一的标签去除、字段规范化等清洗、标准化工作，输出供给下游健康热点识别与处理模块使用，该模块数据处理流程见图 2。

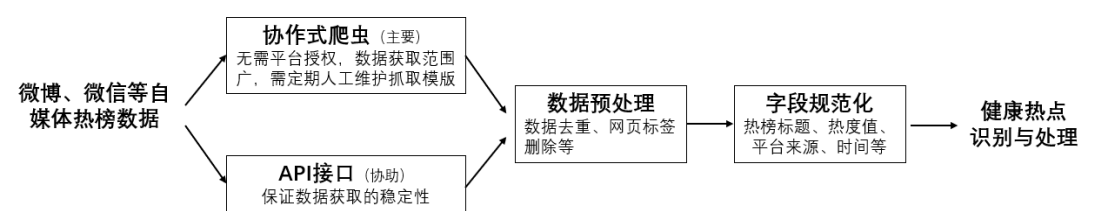


图 2 新媒体热点数据集获取与处理模块流程图

3.3.2 健康热点识别与处理

社交媒体热榜数据中包含大量娱乐、时事、健康等多类型短文本数据，该模块通过前期研究的 ERNIE-Bi-LSTM 模型^[19]，可实时从大量的社交媒体热榜数据中，自动完成健康类热榜数据筛选及信息分类，识别精确度、召回率满足该模块需求。但由于健康热榜短文本数据具有文本内容简短、数据噪声大、特征维度稀疏等特点，使用预训练模型很难完成健康短文本数据的内容拓展与标签丰富化处理。随着人工智能大模型在文本处理方面的能力不断提升，人工智能模型可以通过较少的信息输入和轮对话，即可将短文本快速拓展细节信息，这为健康热点短文本内容批量拓展提供了保障。健康热点识别与处理模块基于本地部署的人工智能模型，将上游健康热点数据输入模型中，模型根据参考的样例数据，完成健康热点信息内容拓展，为医学证据遴选模块提供数据保障，表 1 为人工智能模型对健康热点数据标题改写示例。

表 1 人工智能模型对健康热点数据标题改写示例

原始健康热点标题	人工智能健康热点标题改写
少年每天 5 瓶碳酸饮料患糖尿病	过量糖分摄入导致青少年糖尿病风险增加的研究分析
医生：黄桃罐头缓解甲流症状有道理	维生素 C 对缓解甲型流感症状的辅助作用
反季蔬菜不安全没营养？专家辟谣	反季节蔬菜的安全性营养价值

3.3.3 基于循证医学的证据遴选模块

该模块以人工智能模型为工具，基于近年优质生物医学文献摘要数据，通过综合文献研究类型、研究方法证据等级、发表时间、期刊影响因子、引用频次、主题相关度等因素建立健康热点数据与生物医学文献的动态关联，实现健康热点与学术证据的智能匹配，具

10.12201/bmr.202610.00043V1

体步骤如下：

(1) 构建证据索引库 (IndexDocuments)

利用人工智能模型对入组的生物医学文献数据进行自动标引抽取，并构建向量库。该步骤通过人工智能模型对每篇文献进行分析，提取关键信息并生成对应的向量表示，最终将这些向量存储到证据向量库中，用于后续的多维度文献智能化分析和关联推荐，该模块数据处理流程见图 3。

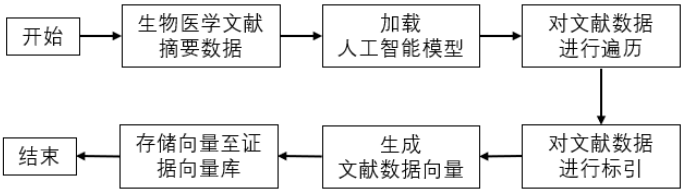


图 3 证据构建索引库模块数据处理流程图

(2) 证据遴选 (SelectEvidence)

从证据向量中遴选与指定健康热点最相关的证据。该步骤通过人工智能模型计算某个健康热点数据的特征维度、分类与证据向量库数据的相关性，返回相关性最高的 10 条证据，输出至下游模块，该模块数据处理流程见图 4。

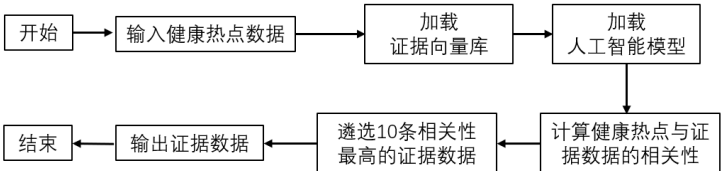


图 4 热点证据关联模块数据处理流程图

(3) 证据总结 (SummarizeEvidence)

该步骤主要对证据遴选模块输入的 10 条证据进行自动概要和观点提取，通过人工智能模型分析每条证据，生成简洁的概要和关键观点，并在概要内容中标注证据出处。最终返回一个包含概要和观点的列表，该模块数据处理流程见图 5。

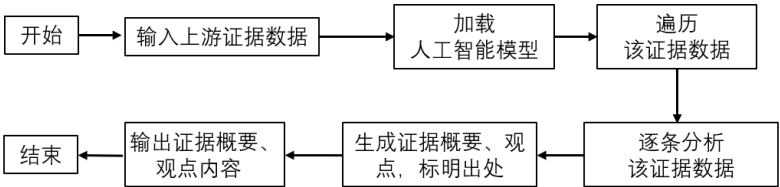


图 5 证据总结模块数据处理流程图

(4) 证据分析 (AnalyzeEvidence)

该步骤主要对证据遴选模块输入的 10 条证据进行多维度分析。首先通过人工智能模型判断其调性（证据数据对健康热点的支持程度），并生成证据概述、研究方法、数据规模等维度的内容。最终返回一个包含分析结果的对象 AnalysisResult，该模块数据处理流程见图 6。

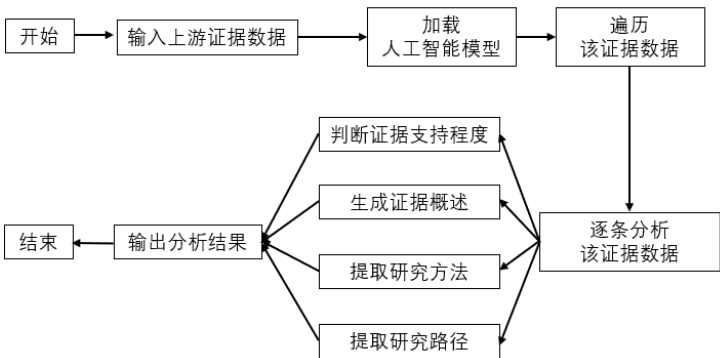


图 6 证据总结模块数据处理流程图

最终基于循证医学的证据遴选模块汇聚以上证据总结、分析等数据，针对健康热点信息提供多维度证据关联。

10.12201/bmr.202610.00043V1

3.3.4 基于人工智能的健康科普内容生成

为避免因使用通用人工智能训练语料库而导致的生成内容“幻觉”，并确保健康科普内容生成的准确性、权威性，该模块以生物医学文献数据中的科普文章、健康热点数据、中国公众健康网科普数据、北京科协云平台数据为底层学习语料，针对微博、微信、小红书等自媒体平台科普文章内容传播特征，迭代训练本地模型。通过输入关键词、需求等提示信息，人工智能模型会根据提示生成健康科普文本、脚本数据。该模块还可根据选定的脚本内容，利用人工智能模型图像生成技术，生成多类型风格的科普配图，通过多轮对话完成科普配图细节调优，满足多场景下的健康科普内容制作。

4 应用效果

基于人工智能驱动的循证健康信息传播系统自 2026 年 1 月正式上线以来，面向健康科普创作人员、科普媒体记者、健康科普传播领域科研人员等多类型用户，提供健康热点查询、健康科普呈现、多维健康证据遴选、健康流言发现、健康科普创作等智能服务，有效提升了健康信息传播的科学性、时效性与公信力。

4.1 打通学术资源科普转化路径，提升学术成果社会化利用效率

系统构建社会健康热点映射学术资源模块，通过数据接口的调用，该模块文献详情页对接生物医学文献智能服务平台^[20]。用户可通过“健康热点”、“循证证据”、“文献获取”的路径关联，实现了从公众健康热点到前沿科研成果的快速转化，有效弥合了公众认知与学术知识之间的鸿沟。其次系统构建基于本地学术资源、健康科普库的语料库，辅助人工智能模型的训练，在健康科普内容生成过程中，从机制上规避通用大模型数据不实风险。从而推动学术资源高效转化为优质科普内容，提升科技成果的传播效率和社会影响力。

4.2 构建循证健康流言识别模块，探索健康流言发现新路径

系统以健康热点识别模块及健康热点与证据遴选模块的动态关联模型为基础，将社交媒体健康热点与高质量学术证据进行快速匹配验证，由模型计算学术证据对健康热点的支持程度，将支持程度低于 50% 的数据标记为疑似健康流言信息，交由人工进一步判断，判断无误后纳入健康流言金标库，用于持续训练健康流言识别模型。该模块为健康流言发现到可核验依据生成提供了流程化支撑，为健康谣言的科学辟谣提供高效、可溯源的技术路径。目前该模块已应用于中国科协科普中国平台“科学辟谣”、北京科技记者协会“科学流言榜”的信息监测中，模块应用见图 7。



图 7 健康热点与证据观点验证图

4.3 构建智能辅助工具体系，赋能健康科普规范化生产

系统面向临床医务人员、医学科研人员、媒体编辑等健康科普创作群体，提供健康热点查询、多维证据遴选、科普图文与视频脚本智能辅助生成等工具支撑，有效降低了专业科普内容生产的门槛与成本，推动健康内容生产向循证化、规范化方向转型升级。系统成果与工具体系已初步具备在医疗机构、公共卫生机构、科协系统及相关媒体平台推广应用的

条件，为健康科普传播体系的智能化升级提供了可复制的示范路径。

随着系统各模块的应用与系统推广，用户对该系统的日均访问量逐步提升，用户关注度显著提升。用户跳出率则仍然维持在 87%，需进一步优化。因此本系统下一步工作重点，将进一步提高系统内容质量，增加系统个性化和推荐模块，以扩大用户规模并降低用户跳出率。

5 结语

本研究提出了一种基于人工智能模型的循证健康信息传播系统。通过建立健康话题与生物医学文献的动态关联模型，实现健康热点与学术证据的智能匹配，为公共健康决策与科普创作提供坚实的信息支撑。本系统进一步尝试实现社交媒体文本、学术文献与健康科普的深度融合，构建跨域健康知识体系。通过语义关联分析与内容融合模型，挖掘公众健康关注点与生物医学数据之间契合点，促进科研成果的应用转化与社会传播。下一步该系统将根据算力硬件规模，分阶段邀请各行业用户开展平台功能测试，努力为科研人员、健康科普从业者提供高效、可信、系统的创作辅助工具。

作者贡献：侯震负责研究设计、论文撰写、论文修订；童惟依、邓靖飞负责健康热点、健康科普数据清洗、遴选；李扬负责论文指导。

利益声明：所有作者均声明不存在利益冲突。

参考文献

10.12201/bmr.202610.00043V1

- 1 黄梦洁,曾云霄,葛蒲,等.社区居民健康科普需求及其影响因素研究[J].中国全科医学,2023,26(04):426-433.
- 2 文怡力,黄友良,王巍洁,等.新媒体环境下我国居民健康素养提升路径研究[J].医学信息学杂志,2024,45(06):63-67.
- 3 祖光怀.循证医学和循证健康教育[J].中国健康教育,2002,(04):59-61
- 4 梁新,李金兰,徐帆.人工智能技术在医务人员健康科普制作中的应用[J].中国信息界,2025,(03):149-151.
- 5 潘越,贾淑娴,陈翔梧,等.循证科普: 如何保证健康科普创作的科学性[J].科普研究,2019,14(05):102-107+114
- 6 易凯谕,韩锡斌. 从混合教学到人智协同教学:生成式人工智能技术变革下的教学新形态[J]. 中国远程教育,2025,45(4):85-98
- 7 Ian J. Goodfellow, Jean Pouget-Abadie, et al. Generative adversarial nets. In Proceedings of the 28th International Conference[C]Cambridge, MA, USA: MIT Press, 2014:2672–2680.
- 8 Ashish Vaswani, Noam Shazeer, et al. Attention Is All You Need[J].arXiv preprint arXiv: 1706.03762v1,2017
- 9 OpenAI.DALL·E Detection Classifier Access Program[EB/OL]. <https://openai.smapply.org/>,2021-01-05
- 10 Higgins, Julian P T et al.The Cochrane Collaboration's tool for assessing risk of bias in randomised trials[J].BMJ (Clinical research ed.),2011,(343):d5928
- 11 李海君,李昀珊,吴一波.辟谣科普教育对健康科普创作者循证科普的影响[J].科技传播,2021,13(02):21-23+76
- 12 权辰昕,刘今,魏淑婷,等.控烟健康传播核心信息循证指南制定方案[J].中国全科医学,2021,24(31):3923-3928.
- 13 米锋.健康科普现状及提升策略[J].中国健康教育,2025,41(06):569-574
- 14 梁新,李金兰,徐帆.人工智能技术在医务人员健康科普制作中的应用[J].中国信息界,2025,(03):149-151.
- 15 王硕,阎妍,李正风.生成式人工智能赋能科学普及: 技术机遇、伦理风险与应对策略[J].科普研究,2024,19(04):5-13+22+101
- 16 陈娟,全思遐.智能化健康科普如何言之有效——基于健康传播理论的 DeepSeek 生成内容评估[J].新闻与写作,2025,(10):58-71.
- 17 张静,江永丰,黄俊杰.国产大语言模型在健康科普文案生成中的对比研究——以数字人播报为应用场景[J].无线互联科技,2025,22(19):80-85.
- 18 童博,田晨,李亿梦,等. 大语言模型辅助循证健康科普作品制作的可行性验证 [J]. 协和医学杂志,2026,17(4):1063-1072
- 19 侯震,童惟依,邓靖飞,等.融合深度学习模型和上下文特征的健康话题短文本分类[J].电视技术,2023,47(07):18-23+27
- 20 中国医学科学院医学信息研究所.生物医学文献智能服务平台 [EB/OL].
<https://www.pubbiomed.ac.cn/home>,2025-12-10