基于知识增强的网络健康信息多维评估智能体研究

熊建英 熊科云 杨 洋

（江西中医药大学智能医学与信息工程学院 南昌 330003）

〔摘要〕 **目的/意义** 设计融合领域知识库与用户特征的智能体方案，以提升网络健康信息评估准确性及决策支持有效性。**方法/过程** 构建医学领域知识库，结合检索增强等方法，实现“多模态解析－知识检索增强－四维评估（可信解释、风险提示、证据引用、个性化建议）”的闭环评估流程。**结果/结论** 该框架可提升健康信息可读性，并辅助支持信息决策。采用基于证据的内容风险评估方法，有助于公众健康信息素养的提升。

〔关键词〕 健康信息素养；知识增强；多维评估；智能体

〔中图分类号〕 R－058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673－6036.2025.11.002

A Multi－dimensional Evaluation Agent for Network Health Information Based on Knowledge Enhancement

XIONG Jianying, XIONG Keyun, YANG Yang

School of Intelligent Medicine and Information Engineering, Jiangxi University of Chinese Medicine, Nanchang 330003, China

〔**Abstract**〕 **Purpose/Significance** To design an agent scheme that integrates domain knowledge bases and user characteristics, so as to enhance the accuracy of network health information assessment and the effectiveness of decision support. **Method/Process** A knowledge base in the medical field is constructed. By integrating methods such as retrieval enhancement and so on, a closed－loop assessment process of “multimodal analysis, knowledge enhancement retrieval, four－dimensional evaluation model (trustworthy interpretation, risk warning, evidence citation, personalized suggestions)” is realized. **Result/Conclusion** This framework can enhance the readability of health information and assist in supporting information－based decision－making. The adoption of evidence－based content risk assessment methods is conducive to the improvement of public health information literacy.

〔**Keywords**〕 health information literacy; knowledge enhancement; multidimensional evaluation; agent

1 引言

互联网中各类健康养生信息的传播，影响着人们的健康观念和行为决策。然而，大量虚假或低质养生信息伪装成“科普”形式，通过夸大疗效、隐匿风险等话术误导用户，甚至延误治疗，形成公共健康风险[1]。《全民健康素养提升三年行动方案

〔修回日期〕 2025－11－14

〔作者简介〕 熊建英，博士，副教授，发表论文 10 余篇。

〔基金项目〕 江西省高校人文社会科学规划项目（项目编号：TQ24103）。

(2024—2027 年)》明确提出要加强健康内容监管，并提升公众健康素养[2]。

目前，针对不良健康信息的治理主要依赖 3 种手段。一是静态辟谣平台。有研究[3]表明，官方辟谣信息公开越早，中立者和轻信者逆转速度越快，网络谣言传播规模越小。但辟谣平台主要依靠人工更新，难以应对海量动态信息。以中国疾病预防控制中心辟谣平台为例，其 2024 年受理的网民举报谣言已多达 1.47 万条。二是基于机器学习的分类模型。该方法在文本分类方面效果较好，但仅能输出二分类结果，缺乏对信息背后医学逻辑的批判性解释，也无法适配用户个体健康状况。三是基于大语言模型（large language model，LLM）的语义理解技术。LLM 拥有强大的语义理解与生成能力，但在缺乏领域知识嵌入的情况下，容易产生“幻觉”，输出错误内容。此外，其生成的普适性输出有时难以理解，尤其对于老年人而言，理解与利用门槛较高[4]。

网络健康信息良莠不齐，通过简单分类评估无法满足公众探寻信息背后因果关系的认知需求，现有方法在评估维度、证据支持与个性化适配方面存在一定局限。因此，本研究提出融合医学知识证据与用户特征的健康信息评估智能体框架，利用其模拟专家评估行为，将健康信息鉴别从表层语义判断升级为证据支持的风险评估，不仅可以为公众提供实时可靠的信息过滤工具，也可为健康信息治理提供新思路。

2 相关工作

2.1 网络健康信息评估的传统技术路径

网络虚假信息识别研究主要依赖自然语言处理技术，主流方法如下。一是基于规则的方法。通过关键词黑名单进行判断，如“根治”“秘方”等词。然而，虚假健康信息往往伪装成“科普”形式，通过敏感词同义替换等方式规避检测，削弱了该方法的泛化能力[5]。二是基于机器学习和深度学习的方法，即通过学习已有虚假信息的特征，将健康信息鉴别转化为分类问题[6]。早期该方法对内容的挖掘依靠人工特征工程，例如，LightGBM 模型[7]融合情感极性、句法复杂度、内容特征、发布者特征等多维特征；也有研究[8]通过构建健康信息画像，从发布者、内容和接收者 3 个维度提取特征。随着深度学习算法的发展，双向编码器表征（bidirectional encoder representations from transformers，BERT）等预训练模型可自动提取虚假信息的主要表述和语义特征，识别虚假信息[9]，一些 BERT 衍生模型（如 BioBERT、PubMedBERT）在医疗文本分类中取得了较好的效果[10]。但是将健康信息评估简化为“语言分类”问题，仅输出分类标签，而不提供批判性解释，难以满足用户的决策需求。

2.2 从“分类”到“评估”的转变

健康信息关乎用户的身心健康与决策风险，其核心并非简单的“真”或“假”二元判断。LLM 的出现，为健康信息评估提供了新的思路和方法。与传统方法不同，LLM 可以理解文本中健康建议与科学证据之间的逻辑关联、分析文本所传达的情感倾向，从而更全面地评估信息的质量和可信度。在生成结果时，可按要求生成多种可能的解释、观点和依据，避免单一视角导致的误判，有助于用户更好地理解评估过程和结论的合理性[11]。然而，LLM 普遍存在“幻觉”问题。为此，研究者提出多种方案，例如，基于多个 LLM 的群体智能方法[12]、将复杂任务分解后再用 LLM 进行分层验证[13]等，以提升虚假信息识别的鲁棒性。检索增强生成技术与外部知识库也可显著降低生成“幻觉”或不准确回应的可能性，如采集权威平台发布的公开健康信息数据，构建知识图谱作为外部知识[14]。但是，LLM 技术在应用中仍有待进一步完善。一方面，通用检索增强技术未区分不同来源知识，而证据类型直接影响信息的可信度。如果未建立“信息主张－证据依据”映射关系，评估可能会偏离医学共识；如果未考虑证据源的权威优先级，可能会引入低质量参考，影响结论的可靠性。另一方面，现有方法大多忽视了用户个体差异，普适性结论可能对特定人群造成误导[5]。例如，宣称“姜黄素抗癌”的内容对普通人群可能属低风险，但对凝血功能障碍患者则存在出血风险，而现有工具无法识别此类差异。

健康信息识别旨在赋能用户决策，其核心不只

在于技术实现，更关乎对用户认知、体验与情感的综合考量[15]。智能体技术为实现上述目标提供了可行路径。该技术以 LLM 为核心，具有规划、工具调用与记忆等能力，可自主调度知识检索、风险推理与用户交互等模块，从构建立动态、可解释的评估流程。因此，本研究拟构建标注证据等级的医学知识库，实现领域知识结构化组织与利用；设计融合多模态解析、知识检索增强与四维评估反馈的闭环验证机制；利用智能体技术实现工具原型，以展示多维评估模式的应用价值。

3 健康信息评估智能体框架构建

3.1 模块构成

智能体以 LLM 为核心引擎，包括观察、思考、行动、记忆模块，见图 1。LLM 核心引擎可完成信息理解、推理、生成和对话，用于协调其他模块工作，处理用户输入的信息，驱动思考模块进行分析，最终生成判断、解释和风险提示的反馈。观察模块用于接收用户待鉴别的信息与用户需求信息。在该模块中，用户可通过粘贴的文本、分享的链接、上传的图片或视频提交要鉴别的信息。智能体也可通过会话引导，获取用户年龄、性别、基础疾病、过敏史、当前症状、用药情况、生活习惯等。思考模块主要完成信息处理、分析、推理等核心环节。如信息主张解析、知识的检索与增强、匹配性评估、不合理成分区分，以及结合用户信息进行个性化风险评估等。行动模块通过调用工具、插件等方式，执行具体任务，支持思考模块，并响应用户需求。如查找证据时，调用知识库查询接口和互联网搜索引擎进行检索增强；在对话交互中，通过结构化对话提示词，引导用户提供要鉴别的信息，以及用户个人信息，并明确告知其信息使用及隐私保护措施；在生成结果时，按提示词生成结构化评估报告。记忆模块存储用户关键信息，提升对话效率和个性化体验。存储的内容包括用户对话历史（可帮助理解上下文）、用户个人信息档案（可避免重复询问），以及知识库更新记录、智能体逻辑规则等，以便不断优化智能体。

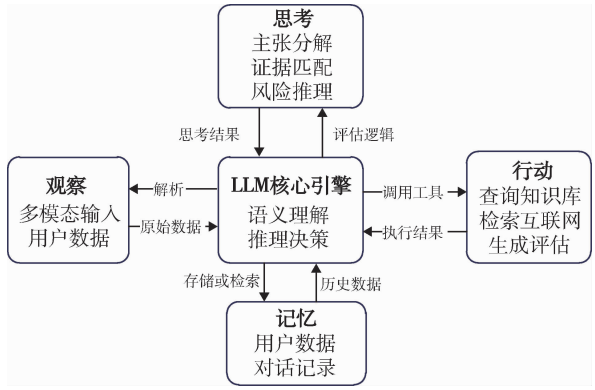

图 1 健康信息评估智能体构成

3.2 智能体工作机制（图 2）

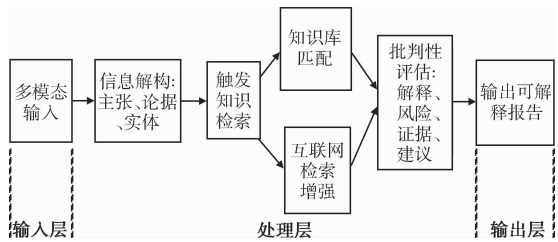

图 2 智能体评估工作流程

3.2.1 输入层 用户可通过微信小程序、互联网、App 等方式接入智能体进行交互。智能体可识别用户输入信息的类型，如果输入是图片或视频，调用图片理解或视频理解类工具接口将其转化为文本供后续分析，如果输入是链接则通过链接读取接口获取信息，系统随后读取记忆模块，检查是否存在该用户个人信息，如果不存在，则启动会话流程引导用户输入。

3.2.2 处理层 智能体通过 LLM 解构用户输入的待验证信息，抽取关键元素，包括主张、证据与实体。构建信息检索条件，采用知识库匹配与互联网检索增强两种方式获取相关证据信息，通过设计提示词，引导 LLM 按照逻辑解释、禁忌风险、证据来源、个性化建议 4 个维度生成融合证据与逻辑的批判性评估，而非简单的价值观判断。

3.2.3 输出层 将处理层生成的内容，进行分层结构化输出。在用户界面可视化展示评估内容，呈现个性化信息评估解释报告。

3.3 领域知识库构建

3.3.1 知识来源 领域知识库是智能体证据支持决策的基础。其构建要遵循权威性、时效性、结构化原则。本研究采用双层知识体系，以医学领域临床指南等权威静态知识为核心层，以互联网中最新研究成果为补充层，主要依托以下几类数据。（1）临床指南。包括权威机构发布的临床指南和标准，如中华医学会发布的临床指南，美国国立临床诊疗指南数据库收录的疾病防治标准等。（2）官方数据。包括药物库、疾病分类、国家卫生健康委员会药食同源目录、世界卫生组织事实清单等。（3）学术文献。包括高质量教科书、经典教材关键章节和权威期刊疾病机制综述等。

3.3.2 知识类型 由于知识库数据来源丰富、类型多样，对不同知识类型采用分类存储策略，见表 1。

表 1 知识分类存储方案

知识类型	表示形式	存储方案	应用场景
实体关系知识	知识图谱	图数据库	禁忌证推理、药物交互检测
文本证据知识	向量嵌入	向量库	语义检索、主张匹配
原始文档知识	结构化 PDF 或 XML	文档库	溯源引用、原文查看

3.3.3 知识库构建 通过知识库集成 3 类核心数据源，见图 3。通过自动化爬虫 + 人工专家审核的方式进行知识更新，爬虫机制通过扫描指南官网或期刊目录，触发增量更新；医学专家组对系统检测到的信息进行定期人工审核，及时更新共识。

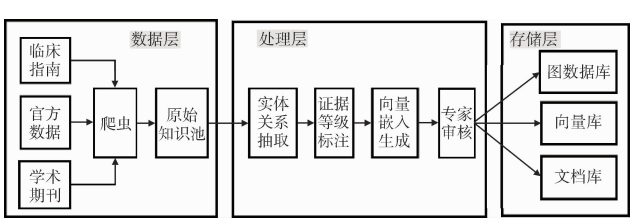

图 3 知识库构建过程

3.4 知识检索增强机制

3.4.1 查询优化机制 利用 LLM 对用户原始查询进行语义解析，将其转化为医学规范表达，如将“血糖高”改为“糖尿病前期”；同义词扩展，如“二甲双胍”改为“Metformin/Glucophage”。该机制统一了用户口语化表达与专业术语，从而提升检索的准确性与效率。

3.4.2 本地知识库优先 将审核过的权威信息结构化存储于本地知识库中，LLM 理解用户查询后，优先在知识库中检索。如果检索到相关证据，返回结构化知识；如果未检索到相关信息，则触发互联网资源检索。

3.4.3 来源分级控制 建立互联网检索优先级：学术数据库 > 政府或权威非营利机构网站 > 三甲医院官网 > 科普平台；并通过黑名单自动屏蔽自媒体营销号和商业推广链接。

3.4.4 结果验证 利用 LLM 提取检索结果的核心论点，与原始查询信息进行相似度计算，确保原始查询与结果摘要的相似度大于阈值。根据检索结果的来源和级别筛选证据，要求证据来自不少于两个独立权威信息源。

4 健康信息评估智能体原型案例实证分析

4.1 智能体评估流程

基于 Coze 平台定义工作流，以实现智能体核心逻辑处理流程。用户界面提供标准化文本输入框，支持用户提交待验证的养生信息内容，并提供查询示例。系统通过对话模式引导用户输入个人健康信息，以便后续提供更精准的分析和建议。

用户输入待验证信息后，智能体通过 LLM 和工作流定义语义解析——>知识检索——>决策分析——>报告生成 4 个步骤。分析报告采用结构化四维评估，包括解释、风险、证据、建议，并采用不同色彩与信息分层方式进行可视化展示。在解释维度，考虑目标用户的信息素养，解释内容兼顾专业解释与通俗化类比方式，见图 4，提高专业术语的可读性，帮助用户理解。在风险预警和证据呈现维度，采用分级提示与具

体说明相结合的方式，见图 5。风险等级为抽象标签，具体说明则符合医疗行业风险沟通规范，可提高用户对风险信息的警惕性；证据呈现中标注等级，并显示证据来源，使证据评估过程透明化。在个性化建议维度，利用提示词引导 LLM 使用包括推荐、禁止、注意 3 项内容的行动导向句式结构，见图 6。通过协同设计，使系统既保持专业严谨性，又根据认知负荷调整信息密度，增强用户对低质量健康信息的批判性理解，提高健康信息素养。

图 4 解释维度可视化展示示例

图 5 风险与证据维度可视化展示示例

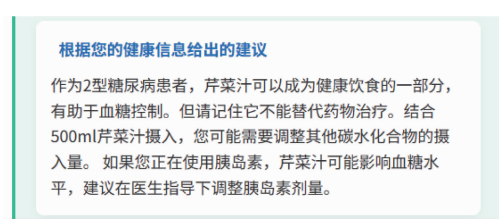

图 6 建议维度可视化展示示例

4.2 结果对比分析

为评估该智能体框架的效果，以网络健康信息“每日饮用大量芹菜汁可以根治糖尿病”为例，模拟其他传统方法的典型输出，见表 2。传统文本分类模型典型输出为“虚假信息”，是一种分类结果。其劣势是无法提供关于“为何虚假”的科学解释，无法区分对普通人群与特定人群的风险差异，无法给出任何证据来源或行动建议。用户知其然不知其所以然。通用 LLM 直接提问后的典型输出为一段概括性文字，可能包含“有一定研究但证据不足”等内容。其劣势是回答缺乏可验证的证据链条，且输出是普适性的，无法实现基于用户画像的个性化风险过滤，存在模型“幻觉”及误导特定人群的风险。

表 2 不同健康信息评估方法效果对比

评估指标	本研究智能体	传统文本分类模型	通用 LLM
评估维度丰富性	四维评估	单一标签	普适性文本解释
证据引用程度	100% 主动提供	不提供	不会主动提供，且部分来源不可考证
可读性	层次分明，易理解	除分类结果外无其他内容	缺乏结构，专业术语多，不易理解
个性化支持	主动	无	非主动

本研究提出的智能体在评估维度、证据来源、结果可读性与个性化支持方面，均显著优于传统分类模型和缺乏领域知识增强的通用 LLM。

5 结语

面对互联网健康信息泛滥与低质化带来的公共健康风险，本研究设计基于大语言模型的智能体框架，将健康信息评估从传统的“文本分类”升级为“医学证据支持下的个性化风险评估”。该框架集成多模态信息解析、知识检索增强与四维评估反馈，构建了从信息输入到可信度验证的闭环流程。原型实践表明，该方法在提升信息可读性、增强决策支持以及改善用户风险认知方面具有明显优势。本研

究目前主要侧重技术框架构建与验证，领域知识库的覆盖广度与更新机制仍有待完善。未来将进一步完善领域知识库，优化检索增强技术与推理算法，以提高评估准确性。

作者贡献：熊建英负责研究设计、论文撰写；熊科云负责文献调研、论文修订；杨洋负责案例设计、论文修订。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 漆晨航. 生成式人工智能的虚假信息风险特征及其治理路径［J］. 情报理论与实践，2024，47（3）：112－120.

2 全民健康素养提升三年行动方案（2024—2027 年）［EB/OL］.［2025－02－17］ https://www.gov.cn/zhengce/zhengceku/202406/content_6955867.htm.

3 王晰巍，李玥琪，邱程程，等. 突发公共卫生事件下网络谣言传播逆转模型及仿真研究［J］. 图书情报工作，2021，65（19）：4－15.

4 刘玮，李泽慧，吴思琪，等. 基于社会网络分析法与 TOPSIS 法的中国老年人健康信息素养困境分析及思考［J］. 中国卫生事业管理，2024，41（11）：1302－1307.

5 王莉. 网络虚假信息检测技术研究与展望［J］. 太原理工大学学报，2022，53（3）：397－404.

6 詹骞，赵冰洁. 健康类虚假信息的人工神经网络识别与治理［J］. 现代传播（中国传媒大学学报），2022，44（8）：155－161.

7 金燕，徐何贤，毕崇武. 多维特征融合的虚假健康信息识别方法研究：基于 LightGBM 算法［J］. 情报理论与实践，2023，46（8）：156－164.

8 赵又霖，庞航远，石燕青. 健康信息画像构建及虚假健康信息识别：融合社会感知数据与发布者先验知识［J］. 图书情报知识，2024，41（6）：141－154，165.

9 冯由玲，康鑫，周金娉，等. 基于 BERT－BiLSTM 混合模型的社交媒体虚假信息识别研究［J］. 情报科学，2024，42（6）：89－98.

10 李盛青，苏前敏，黄继汉. 基于 BioBERT 与 BiLSTM 的临床试验纳排标准命名实体识别［J］. 中国医学物理学杂志，2024，41（1）：125－132.

11 TAN D，HUANG Y，LIU M，et al. Identification of online health information using large pretrained language models：mixed methods study［J］ Journal of medical internet research，2025，27（5）：1－19.

12 何静，沈阳，谢润锋. 大语言模型幻觉现象的分类识别与优化研究［J］. 计算机科学与探索，2025，19（5）：1295－1301.

13 张君冬，刘江峰，邓景鹏，等. 以模治模：基于生成式人工智能的失真健康信息识别［J］. 情报杂志，2025，44（5）：130－138.

14 杨雅娴，吴金红，吴彦坤，等. 融合知识图谱和大语言模型的虚假健康信息识别方法研究［J］. 情报理论与实践，2025，48（3）：127－133.

15 章小童，杨帆，王晓瑜，等. 健康信息搜索系统交互评估理论模型构建研究［J］. 医学信息学杂志，2025，46（4）：1－7，22.

关于《医学信息学杂志》启用“科技期刊学术不端文献检测系统”的启事

为了提高编辑部对于学术不端文献的辨别能力，端正学风，维护作者权益，《医学信息学杂志》已正式启用“科技期刊学术不端文献检测系统”，对来稿进行逐篇检查。该系统以《中国学术文献网络出版总库》为全文比对数据库，可检测抄袭与剽窃、伪造、篡改、不当署名、一稿多投等学术不端文献。如查出作者所投稿件存在上述学术不端行为，本刊将立即做退稿处理并予以警告。希望广大作者在论文撰写中保持严谨、谨慎、端正的态度，自觉抵制任何有损学术声誉的行为。

《医学信息学杂志》编辑部