人工智能引用的健康来源中的权威性信号：评估 ChatGPT 回应中来源可信度的框架

Erin Jacques[1], Erela Datuowei[2], Vincent Jones II[1], Corey Basch[3], Celeta Vanderpool[2], Nkechi Udeozo[4], Griselda Chapa[1]

[1]健康与人类表现系，约克学院，纽约城市大学，94-20 Guy R. Brewer Blvd.，牙买加，纽约州 11451

[2]健康研究与应用教育心理学系，哥伦比亚大学师范学院，525 West 120th Street，纽约，纽约州 10027

[3]公共卫生系，威廉·帕特森大学，300 Pompton Road，韦恩，新泽西州 07470

[4]卫生政策与管理系，纽约城市大学公共卫生学院，55 W 125th St，纽约，纽约州 10027

通讯作者：ejacques1@york.cuny.edu

关键词：
权威性信号、ChatGPT、健康信息质量、答案引擎优化、来源可信度

摘要

自大语言模型（LLM）兴起以来，健康信息寻求的方式发生了根本性变化，ChatGPT 的 8 亿用户中，近三分之一每周都会提出健康问题。了解这些人工智能生成回复的信息来源至关重要，因为健康组织和服务提供者也在投资数字战略，以提升自身在 ChatGPT 等大语言模型系统中的自然排名、触达率和可见度。随着人工智能搜索优化策略日趋成熟，本研究提出了一个权威信号框架，分为四个领域，分别反映健康信息寻求的关键组成部分：首先是“谁撰写了内容？”（作者资历），其次是“谁发布了内容？”（机构隶属关系）、“内容如何经过审核？”（质量保证），以及“人工智能如何找到内容？”（数字权威）。本描述性横断面研究从 HealthSearchQA 中随机选取了 100 个问题。该数据集包含 3,173 个消费者健康问题，由 Google Research 根据公开的搜索引擎建议整理而成。研究人员将这些问题输入 ChatGPT 5.2 Pro，并依据权威信号框架的四个领域记录和编码引用来源。研究对所有引用来源（n=615）进行了描述性统计，并通过交叉列表考察不同组织类型之间的差异。ChatGPT 生成的健康问题回复中，超过 75% 的引用来源来自成熟的机构来源，例如梅奥诊所、克利夫兰诊所、维基百科、英国国家医疗服务体系和 PubMed；其余引用则来自缺乏成熟机构支持的其他健康信息来源。

1. 引言

随着 ChatGPT 的 8 亿用户中近三分之一每周都会提出健康问题（OpenAI，2025年），人工智能在健康指导方面的应用率不断上升；这些用户很可能还会根据回答做出医疗决策。调查显示，高达 78% 的用户承认曾使用 ChatGPT 进行自我诊断（Shahsavar 和 Choudhury，2023年），医疗保健专业人员的使用率也相近。近半数护士表示每周都会使用人工智能（美国医学会，2024年）辅助临床决策（Iqbal等人，2025年），而医生的使用率仅一年间便从 38% 翻倍至 66%（美国医学会，2024年）。鉴于这种广泛应用会影响医疗结果，一个关键问题随之浮现：这些系统实际引用的健康信息来源究竟是什么？

显然，健康信息寻求的格局已经发生变化，这在很大程度上归因于 OpenAI 于 2022 年推出 ChatGPT，并在 2023 年引入实时网络搜索能力后进一步得到增强（Search Engine Journal，2025年）。个人设备让人们能够轻松获取健康信息，加之新冠肺炎疫情后人们对医生和医院的信任大幅下降（Perlis 等，2024年），促使人们转向非传统健康信息来源（Casselman-Hontalas 等，2024年）。消费者转而使用人工智能驱动系统寻求健康信息，这一趋势给这些来源的可信度带来了新的挑战。

ChatGPT 展现出进行诊断和提出建议的能力，在某些情况下甚至优于获得人工智能支持的医生（CADTH，2025年）。然而，人们日益担忧 ChatGPT 所提供医疗信息的效度和可靠性，其中包括公平性、偏见、透明度和隐私等问题（Haltaufderheide 与 Ranisch，2024年）。医疗专业人员和公众都列举了 AI 聊天机器人提供误导性信息和产生幻觉的案例（Tangsrivimol 等，2025年）。

有许多可信度信号可用于识别可信赖的健康信息（Fahy等人，2014年）。表明内容真实可靠的既定标志包括专业资质（医学博士或哲学博士）、与政府机构或学术医疗中心的机构隶属关系、同行评审编辑流程，以及循证引文（MedlinePlus，2024年）。然而，辨别优质健康信息需要具备健康素养，而 88% 的美国成年人健康素养未达到熟练水平（Kutner等人，2006年）。因此，许多消费者无法识别这些既定的可信度信号（Sun等人，2019年；Diviani等人，2015年）。

从网络搜索转向能够即时提供答复的人工智能系统，使得对健康信息进行系统化质量监测变得至关重要。近期研究显示，大语言模型生成的健康答复中，有50%至90%未能得到其引用来源的充分佐证（Wu等人，2025年）。这引发了人们对传统可信度

指标在人工智能系统中指导来源选择的有效性的担忧，尤其是正当的健康组织和另类信息提供者仍在争夺曝光度，而人工智能系统的透明度有限，且没有标准化的排名标准（Chen等人，2024年）。

提高来源在人工智能生成的回答中可见度的商业动机，可能会放大来源的影响力，而不论其医学可信度如何（Chen et al., 2024年）。再加上当前大语言模型系统缺乏引用透明度，随着优化实践日益成熟，回答操纵的可能性也随之增加（Gao et al., 2023年）。鉴于健康错误信息可能导致延迟就医或采用无效治疗等后果，来源审查和排序变得尤为重要（Memon et al., 2022年; Tong et al., 2023年）。

了解人工智能系统实际引用了哪些来源，以及这些来源的特征，对于评估这些系统是否优先考虑既有质量标志，或在多大程度上容易受到操纵，至关重要。作者资历、机构隶属关系和编辑审查流程等权威信号，长期以来一直被视为健康信息质量的基础指标（Silberg et al., 1997年）。确定人工智能引用的健康来源中实际存在哪些权威信号，是了解这些系统是否优先考虑既有质量标志或其他标准的必要第一步。

1.1 研究空白

近期研究从多个角度考察了 ChatGPT 在医疗保健场景中的表现。系统综述评估了 ChatGPT 生成的医学内容的准确性，发现其总体准确率为 56%（Cai 等，2024年）；研究还发现，回答准确性会因领域而异（Ge 等，2025年）。另有研究考察了引文可靠性。Hosseini 等（2023年）发现，ChatGPT 生成的引用中有 47% 是虚假的，93% 不准确。

以往研究主要关注答案准确性和引文可靠性，而非对来源可信度进行整体系统性评估。作者资历、机构隶属关系和编辑审查流程等权威信号，长期以来一直被视为衡量健康信息质量的既定指标（Silberg 等，1997年），但目前尚无框架能够对 AI 引用的健康信息来源中的这些信号进行系统性描述。

围绕如何通过优化提升来源在大语言模型中的可见度，相关术语仍存在广泛争议。业内人士使用的术语包括答案引擎优化（AEO）、生成式引擎优化（GEO）、生成式搜索优化（GSO）和人工智能答案引擎优化（人工智能 AEO），但对这些术语的定义各不相同、细微含义不一，有时还会交替使用。尽管行业对这些术语的细微含义尚未形成共识，但它们都描述了提升内容在大语言模型回答中可见度的策略（Digiday，2025年；Backlinko，2025年）。虽然这些优化技术的标准化命名仍有争议，创业领域已在不断迭代方法，以提高其在人工智能引文模式中的呈现比例。

了解人工智能系统最有可能引用的来源具有哪些特征，对于评估这些系统是否容易受到操纵以及容易受到何种程度的操纵至关重要。本研究旨在趁优化策略尚未广泛普及、仍处于早期实验阶段时，确立基线权威信号模式。要了解人工智能系统优先选择哪些来源，需要提出四个关键问题：

1. 谁撰写了这份内容？（作者资历）
2. 谁发布了这份内容？（机构隶属关系）
3. 这份内容经过了怎样的审核？（质量保证）
4. 人工智能如何发现这份内容？（数字权威）

本研究介绍了权威信号框架，该框架针对以下四个领域系统编码了11项权威性信号，用于刻画人工智能引用的健康来源中存在的可信度标志（见图1）。

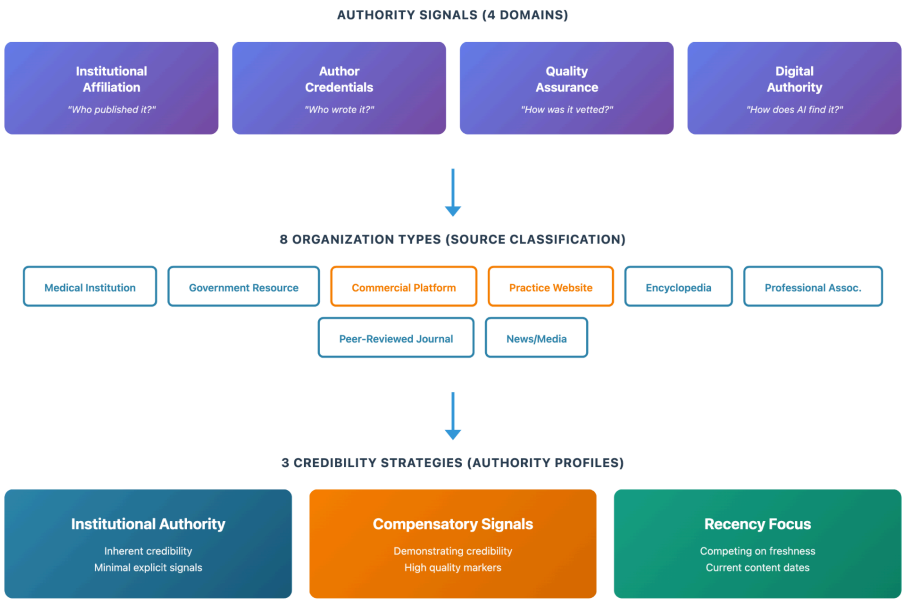


图1：权威信号框架。概念模型展示了权威信号领域、组织类型与可信度策略之间的关系。研究使用四个权威信号领域（机构隶属关系、作者资历、质量保证、数字权威），将 615 个健康信息来源划分为八种组织类型。

权威信号框架通过对以下四个领域中可观察的来源特征进行编码来运作，每个领域都针对一个关键的可信度问题：

作者资历回答“谁撰写了内容？”，通过专业学位、认证和医学专科来体现个人层面的专业能力。变量范围从未注明作者署名，到明确列出医学资质和专科信息。

机构隶属关系回答“由谁发布？”，依据组织类型对来源进行分类，并指出在健康信息语境中，不同的组织背景（政府机构、医疗机构、学术机构、商业平台、专业执业机构）具有不同程度的固有机构可信度。

质量保证回答“内容经过了怎样的审核？”，通过记录医学审查声明、内容时效性以及所引用的支持性参考文献，展示编辑流程和证据标准。

数字权威通过量化网络可信度信号（包括 Moz）来回答“人工智能如何找到它？”

页面权威度、域名权威度、垃圾信息分数、架构标记的实施情况以及内容长度。

每个变量均依据明确的编码标准进行操作化定义，以确保研究结果可复现（可通过 Zenodo 获取：https://doi.org/10.5281/zenodo.18287499）。这种方法承认了影响数字健康信息来源可信度的各类交织信号。

本研究将这一框架应用于 ChatGPT 针对 HealthSearchQA 数据集中的 100 个经验证的消费者健康问题所引用的来源，以确定人工智能系统优先考虑的是公认的质量标志，还是其他选择标准。了解这一点对于理解人工智能在健康信息传播中的作用至关重要。

2. 方法

2.1 数据来源与样本

这项描述性横断面研究使用了 HealthSearchQA 的数据，其中包含 Google Research 从公开可用的搜索引擎建议中整理的 3,173 个消费者健康问题。这些问题反映了与医学搜索相关的常见查询

词。问题经过随机排序，并选取前 100 个纳入研究。

一名编码员使用新创建的 ChatGPT 账户和电子邮箱，将这 100 个健康问题逐一手动输入 ChatGPT 5.2 Pro，以避免先前对话产生影响。所有数据均于2026年1月11日收集，每个问题分别输入独立的聊天会话，以避免后续回答受到先前查询的影响。为确保回答提供来源，每个问题后都附上标准提示语：“请在回答中附上来源链接。”每个问题生成了 4+ 个来源，最终得到 618 条初始回答；Google Sheets 中的每一行代表一个独立来源。由于网站无法访问以及出现 403 和 404 错误，删除了三条回答，最终样本量为 615。另有一名编码员对部分聊天记录进行核查，以确保所有来源均准确转录至电子表格。

2.2 变量与测量

数字权威、质量保证、机构隶属关系和作者资历这四个领域中的权威信号分别归类如下：

1. 机构所属领域指组织类型，包括以下八类，编码为1至8： 1=医疗机构、2=政府资源、 3=商业健康信息、4=专业/执业网站、5=百科全书、6=专业协会、7=同行评审期刊和8=新闻/媒体

2. 作者资历领域反映作者的署名级别，编码为 0=无、1=仅有姓名、2=姓名及资历。这两个作者变量—— [可见作者署名（是/否）；作者资历（是/否）——]随后被汇总为综合的作者署名级别

3. 质量保证领域包括以下四类：引用参考文献（是/否）；医学审查声明可见（是/否）；内容时效性（0=2020年以前或未列出、 1=2020-2023年、2=2024-2026年）。用于确定基于网络的内容由人工智能生成还是由人类生成的原创性评分范围为0至100分。

4. 数字权威域名包含六项指标：页面权威度、域名权威度和垃圾评分，均按 0–100 的尺度计分（= 越高表示垃圾信号越多）；内容长度分为（0=简短、1=适中、2=全面），架构标记（有/无），并编码为 Schema 类型（0=无、 1=微数据、2=JSON-LD）

可通过 Zenodo 获取完整的操作性定义[https://doi.org/10.5281/zenodo.18287499]

2.3 测量程序

自动化编码流程：鉴于数据集规模庞大（615 个来源），且需要一致地应用操作性定义的权威信号，因此开发了自动化编码流程。研究人员使用 Apify 进行网页抓取，系统地收集健康网站内容。研究人员 EJ 在人工智能的协助下编制了编码手册，制定明确的操作性定义，作为使用 Claude API 检查各网站类别并编码详细信息的标准。该流程部署为 Vercel 上的无服务器函数，以便批量处理 URL。结果存储在 Supabase（PostgreSQL数据库）中，并导出至 Google Sheets 进行验证。对于自动化流程无法从 URL 提取内容而导致的缺失数据字段，两位编码人员均进行了审查并手动录入。随后，质量控制环节使用 JavaScript浏览器扩展代码，按照相同的操作性定义验证编码数据并标记缺失字段。该流程成功编码了以下类别：来源名称、组织名称、组织类型、机构隶属关系、作者姓名、作者资历、时间日期、内容长度类别、医学审查声明、引用参考文献、参考文献数量、包含Schema、Schema类型

完整的组织分类映射可通过 Zenodo 获取（https://doi.org/10.5281/zenodo.18287499）

内容时效性编码：通过 Jamovi v2.3 进行频数分析，确定内容时效性，发现日期自然聚集的现象。从自动化流程中提取时间日期，随后按照以下编码规则将其编码到时效性列联列中： 0 =2020年之前或未列出日期； 1 =2020—2023年和 2 = 2024-2026。

API 和批量编码：页面权威度、域名权威度评分（DA 评分）和垃圾评分（0-100；数值越高，= 垃圾信号越多）通过 Moz API 收集。内容原创性使用 Originality.ai 进行评估，

该工具成功对约 50% 的来源进行了分类（n=~309）。其余来源在原创性分析方面出现错误。

人工分类：在自动化数据收集过程中，提取了每个来源的组织名称和机构隶属关系，共得到 248 个唯一组织。通过查看组织名称和机构隶属关系，确定最合适的类别，以此对组织类型进行分类。若分类存在歧义，则使用 DA 评分作为第三项指标来确认组织类型，因为较高的 DA 分数通常与规模较大的运营相对应。在这些情况下，还会进一步审查组织网站。在商业健康信息与专业/执业网站之间，域名权威度作为额外的分类依据：DA ≤ 45 表示专业/执业网站，DA > 45 表示商业健康信息（图 3）。本研究建立了一套包含 8 个类别的分类系统，并按 1–8 编码：(1) 医疗机构、(2) 政府资源、(3) 商业健康信息、(4) 专业/执业网站、(5) 百科全书、(6) 专业协会、(7) 同行评审期刊和 (8) 新闻/媒体。组织类型的完整定义及来源清单见**补充附录A**。

3. 分析

对所有引用来源（n=615）计算了描述性统计，包括频数、百分比、中位数和百分位数。此外，还进行了交叉列表，以考察以下权威性信号在八种组织类型中的分布：医学审核、引用参考文献、架构标记、内容长度和内容时效性。数据准备使用的软件包括 Google Sheets，统计分析使用 Jamovi。

4. 结果

引用次数主要来自具有固有机构权威的来源（75.7%），包括医疗机构、政府资源、专业协会、同行评审期刊、百科全书、新闻和媒体。其余24.3%来自商业健康信息平台（12.4%）和专业/执业网站（11.9%），这些来源不属于传统机构结构。

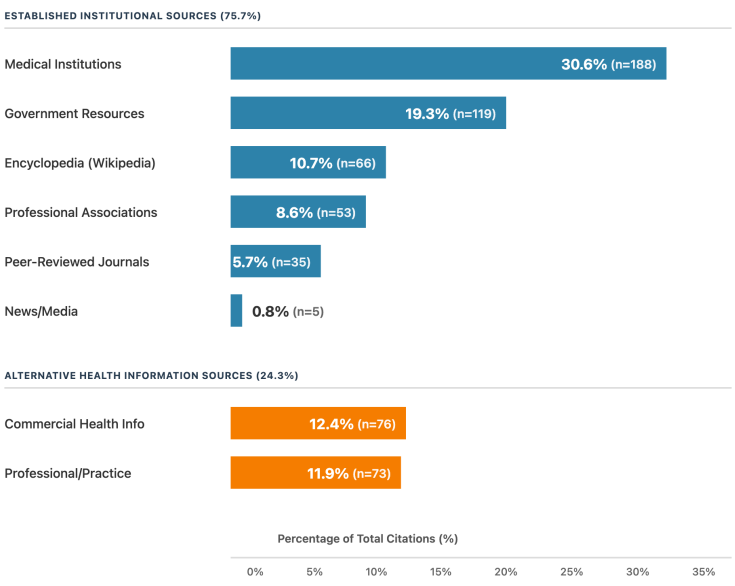


图2： ChatGPT 健康相关引用次数中的机构主导地位：按组织类型划分的 615 个 ChatGPT 引用健康信息来源的分发情况，展示了成熟机构来源与另类健康信息来源之间的区别。各类别的百分比和来源数量（n）如下所示。

进一步细分已确立的机构来源，医疗机构占比最高，为30.6%；其次是政府资源（19.3%）、百科全书（10.7%）、专业协会（8.6%）、同行评审期刊（5.7%），新闻/媒体占比最低（0.8%）。总体而言，引用次数显示出显著的引用集中于特定组织。共有10个组织占所有健康信息引用的52.8%，其中维基百科占比最高，为10.7%；其次是梅奥诊所，占9.9%，克利夫兰诊所占9.8%（图3）。

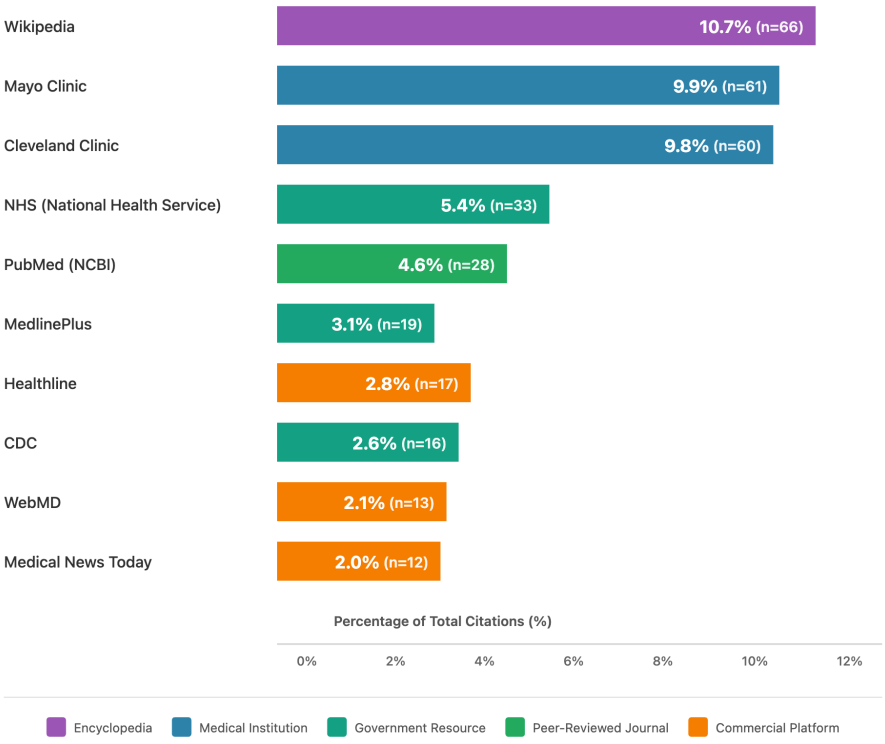


图3：最常被引用的健康信息来源。 *ChatGPT*引用的来源中，排名前*10*的组织占*615*次引用总数中的*325*次（*52.8%*）。颜色表示组织类型：百科全书（紫色）、医疗机构（蓝色）、政府资源（青绿色）、同行评审期刊（绿色）和商业平台（橙色）。

4.1 对四个领域的权威信号进行考察，可见明显规律。

作者资历领域：大多数来源（64.7%）未注明作者署名，28.3%的来源提供了部分署名，只有7.0%的来源完整注明作者署名。

质量保证领域：被引用来源的质量保证机制实施情况参差不齐

来源。只有39.2%的来源列出了支持其健康内容的参考文献，因此大多数来源（60.8%）没有引用参考文献。此外，70.4%的来源没有注明任何医学审查声明，许多来源的内容也未及时更新：40.7%的来源发布日期早于2020年或没有标注日期；其余来源中，22.9%发表于2020年至2023年，36.4%发表于2024年至2026年。另使用Originality.ai进行了单独分析，以检测人工智能生成内容。该工具成功

对约50%的来源进行了分类（n=305）。在305个已分类的来源中，22.6%被判定为人工智能生成，77.4%被判定为人工生成。

数字权威域名：ChatGPT 引用的来源，其页面权威分数中位数为 55，域名权威分数中位数为 89，垃圾信息分数较低，为 2。样本中，47.3% 的内容长度达到全面程度，其余内容中，全面程度中等的占 33.7%，简短内容占 18.5%。多数来源采用了架构标记和 JSON-LD 格式，比例分别为 74.3% 和 68.3%。

4.2 不同组织类型的权威信号特征

四类分类信号标记——引用参考文献、架构标记、内容长度和内容时效性——在不同组织类型中的权威信号实施情况差异显著。医学审查声明在商业健康信息平台（71.1%）和医疗机构（49.5%）中最为常见，而政府资源（13.4%）、同行评审期刊（11.4%）、专业/执业网站（4.1%）和百科全书（1.5%）中的占比则较低。

在考察引用参考文献时，百科全书（100.0%）和同行评审期刊（80.0%）大量提供了来源引文，相比之下，医疗机构（13.3%）和专业/执业网站（19.2%）提供的比例较低。

不同组织类型采用架构标记的情况总体一致，其中百科全书（100.0%）和新闻/媒体机构（100.0%）占比最高，其次是商业平台（86.8%）和专业/执业网站（78.1%）；然而，同行评审期刊采用架构标记的比例相对较低（42.9%）。

网页文本的内容长度显示，百科全书（83.3%）和商业健康平台（68.4%）经常提供全面内容（>1,500 字）。专业/执业网站提供全面内容的比例为 50.7%；医疗机构（31.9%）和政府资源（31.1%）提供全面内容的比例最低，在简短、中等和全面内容之间分布更为均衡。

根据网页上列出的日期衡量内容的新近程度，专业/执业网站中近期更新或标注更新日期的比例最高（2024—2026年发布的占61.6%），其次是政府资源（50.4%）和医疗机构（41.0%）。商业平台（18.4%）和同行评审期刊（2.9%）标注近期日期的比例较低。

5. 讨论

5.1 AI 引用的健康信息来源中的三种可信度策略

ChatGPT 引用的来源中，超过 75% 来自成熟的机构来源（医疗机构、政府机构、同行评审期刊、专业协会、新闻媒体和维基百科）。其余来源则具有各种数字权威标记，这些标记可能促使 ChatGPT 将其选为来源。

这些发现揭示了三种截然不同的可信度策略（图 4），或可为组织带来竞争差异化。首先，成熟的机构来源依赖内在权威，而非其他健康信息来源所采用的明确信号。其次，商业平台注重多个领域的质量标志。第三，专业/执业网站保持着较高的内容时效性。

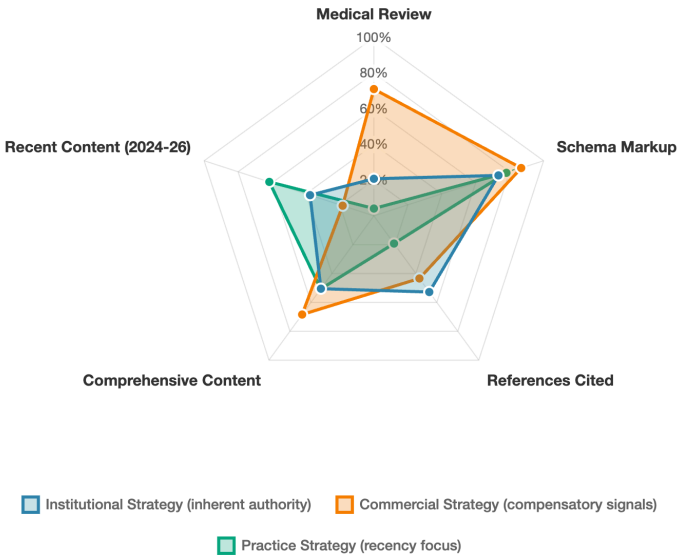


图4：AI 引用的健康信息来源中的三种可信度策略。此图展示了不同权威信号的差异：蓝色代表具有成熟制度的机构来源，橙色代表商业平台，绿色代表专业/执业网站。

首先，机构来源包括政府机构、医疗机构和同行评审期刊，它们都受益于内在可信度优势，因为这些来源依托既有的标准化审查机制（Kington等人，2021）。这种机构权威似乎降低了人们对明确质量保证信号的需求：政府资源仅有13.4%的案例注明经过医学审核，同行评审期刊的这一比例也仅为11.4%；而这两类组织采用架构标记的比例均低于样本中位数（分别为61.3%和42.9%）。

超过半数来源的 DA 分数高于 88，但这并非商业网站和专业执业网站表现强劲的原因。相反，其他权威信号的组合可能影响了它们在 ChatGPT 中的可见度。缺乏机构

支撑作用，体现出在多个领域对补偿性可信度信号的投入。商业平台在 71.1% 的来源中注明经过医学审核，在 86.8% 的来源中采用架构标记，并在 68.4 % 的案例中提供全面内容（>1500 字），这三项指标的比例均显著高于机构来源。相比之下，专业执业网站中有更高比例（61.6%）的内容标注了最近两年的日期，体现为

时效日期。其他研究人员也证实了这一发现，他们指出，大语言模型更偏好日期较新的内容（Fang等人，2025年）。

对于商业网站和专业执业网站而言，通过架构标记、质量保证披露（医学审查声明）、内容深度和近期日期来实施技术优化，是良好实践的一种体现；然而，这也可能被用作一种刻意策略，以弥补机构可信度的缺失，并以此影响人类感知和 AI算法检索选择。

权威信号框架考察 ChatGPT 的来源引文，探讨从“由谁发布”到“人工智能如何找到它”的过程。结果显示，组织类型以既有医疗机构为主，而作者身份、编辑流程和技术优化等因素则是次要信号。这些结果对于健康组织和专业人员如何应对不可避免的人工智能搜索优化可见性时代具有重要意义。

5.2 对健康信息质量与来源监测的启示

随着这类应用日益普及，OpenAI 已宣布推出 ChatGPT Health，但目前仍通过候补名单逐步开放。该服务将允许用户输入医疗信息，以获得更个性化的医疗指导（OpenAI，2026年）。鉴于人工智能健康信息检索十分普遍，本研究确立了基线引文模式，以便进行长期追踪。本文记录的权威性信号模式直接关系到患者安全。健康错误信息可能造成严重后果，包括延迟就医以及采用无效且危险的治疗（Memon 等，2022；Tong 等，2023）。既有机构健康信息来源和健康专业人员切身关注人工智能引文规范，以免在可见性方面拱手让位于采用补偿性信号的商业实体。仍有必要持续考察 ChatGPT 引文模式，因为人工智能搜索优化策略通常被称为 AEO、GEO、GSO、AI SEO 及其他变体。

6. 局限性

本研究受到横断面设计性质的限制，所有数据提取均在同一时间点进行，因此无法了解这些数据会如何随时间变化。鉴于本研究使用了 HealthSearchQA数据集中的一小部分来考察这些引用模式，因此尚不清楚，若使用数据集中另一部分问题或一组完全不同的问题，结果是否仍会保持一致。

7. 结论

本研究提出了一个权威信号框架，分为四个领域，反映了健康信息寻求的关键组成部分：首先是谁撰写了内容（作者资历），其次是谁发布了内容（机构隶属关系），再者内容经过了怎样的审核（质量保证），最后人工智能如何发现内容（数字权威）。ChatGPT 生成的健康信息回复中引用的主要来源是知名机构来源。不到四分之一的引用来自商业及专业实践类网页，这些网页所属的组织类型是次要影响因素。本项 ChatGPT 分析确立了人工智能搜索优化策略的当前基准模式，早于缺乏成熟机构支持的健康信息机构广泛实施优化策略。本研究探讨了人工智能健康信息引文行为，为 2026年伊始的人工智能搜索优化策略提供了参考，早于其游戏化，并有助于随着相关策略的发展进行纵向监测和比较。

数据可用性

所有研究项目 m　材料和编码流程可在以下网址公开获取：https://doi.org/10.5281/zenodo.1828749　　9

人工智能声明

人工智能用于支持数据收集、文献检索、整体格式编排、图像和表格制作，以及语法和句子表达方面的编辑。

软件与工具

Moz.（2024年）。*Moz* 域名权威度。Moz Pro。检索自 https://moz.com/learn/seo/domain-authorityOriginality.AI。（2024年）。*AI* 内容检测工具（2.0 版）[软件]。检索自 https://originality.a Anthropic。（2024年）。*Claude API*：*Claude Sonnet4* [应用程序接口]。检索自 https://www.anthropic.com/api Apify Technologies。（2024年）。网站内容爬虫 [网页抓取工具]。检索自 https://apify.com/apify/website-content-crawler Google。（2024 年）。*Google Sheets API*（V4）[应用程序接口]。https://developers.google.com/sheets/api Supabase Inc.（2024年）。*Supabase*：开源 *Firebase* 替代方案 [软件]。检索自 https://supabase.com

参考文献

美国医学会。（2024年）。医生在临床实践中采用人工智能。https://www.ama-assn.org

美国医学会。（2024年）。美国医学会增强智能研究：医生对在医疗保健中使用人工智能的态度。检索自 https://www.ama-assn.org/practice-management/digital-health/2-3-physicians-are-using-health-ai-78-2023

Anthropic。（2024年）。*Claude API*: *Claude Sonnet 4*[应用程序接口]。https://www.anthropic.com/api

Apify Technologies。（2024年）。网站内容爬虫[网页抓取工具]。https://apify.com/apify/website-content-crawler

Cai, Z. R. 等。（2024年）。对ChatGPT生成的医疗回答的评估：系统综述与荟萃分析。《生物医学信息学杂志》，*150*，104588。

加拿大卫生药品和技术局（CADTH）。（2025年）。*2025*年观察清单：医疗保健中的人工智能。NCBI Bookshelf。https://www.ncbi.nlm.nih.gov/books/NBK613808/

Casselman-Hontalas, K. 等（2024年）。COVID-19疫情期间健康信息来源的信任变化。《健康传播杂志》。

Chen, J. 等（2024年）。健康信息系统中的生成式引擎优化与人工智能可见度。《数字健康杂志》。

Diviani, N.、van den Putte, B.、Giani, S. 和 van Weert, J. C. M.（2015年）。健康素养较低与在线健康信息评估：文献系统综述。《Journal of Medical Internet Research》，17(5)，e112。https://doi.org/10.2196/jmir.4018

Fahy, E. 等（2014年）。健康信息中的可信度信号：系统综述。《健康信息与图书馆杂志》。

Fang, X.、Wang, H. 和 Sakai, T.（2025年）。大语言模型是否偏好近期内容？关于基于大语言模型的重排序中的近因偏差的研究。早稻田大学研究出版物。

Gao, L. 等（2023年）。大语言模型系统中的引用透明度。《自然·机器智能》。

Ge, J. 等（2025年）。评估 ChatGPT 对健康相关问题的可靠性：系统综述。《信息》，12(1)，9。

Google。（2024年）。*Google Sheets API*（第 4 版）[应用程序接口]。https://developers.google.com/sheets/api

Haltaufderheide, J., & Ranisch, R.（2024年）。关于 ChatGPT 医疗信息效度和可靠性的伦理关切。《医学伦理学杂志》。

Hosseini, M., Gao, C. A., 等（2023年）。ChatGPT 生成的医疗内容中伪造和不准确参考文献的比例很高。*Cureus*。

Iqbal, S., 等（2025年）。在医疗环境中使用 ChatGPT 进行临床决策和诊断。《临床医学杂志》。

Jacques, E.（2026年）。人工智能引用的健康信息来源中的权威性信号：自动化研究流程（第 1.0.0 版）[计算机软件]。Zenodo。https://doi.org/10.5281/zenodo.18287499

Kington, R., Arnesen, S., Chou, W. S., Curry, S. J., Lazer, D., & Villarruel, A.（2021）。在社交媒体中识别可信的健康信息来源：原则与属性。NAM Perspectives。讨论论文，美国国家医学院。https://doi.org/10.31478/202107b

Kutner, M. 等（2006年）。美国成年人的健康素养：*2003*年全国成人读写能力评估结果。美国教育部，国家教育统计中心。

MedlinePlus。（2024年）。评估健康信息。美国国家医学图书馆。https://medlineplus.gov/evaluatinghealthinformation.html

Memon, S. A. 等人（2022）。健康错误信息的后果：延迟就医和采用无效治疗。《英国医学杂志全球健康》。

Moz。（2024年）。*Moz* 域名权威度。Moz Pro。https://moz.com/learn/seo/domain-authority

OpenAI.（2025年）。*ChatGPT* 使用统计与健康信息寻求。https://openai.com

OpenAI.（2026年）。推出 *ChatGPT Health*。https://openai.com/index/introducing-chatgpt-health/

Originality.AI.（2024年）。*AI* 内容检测工具（2.0 版）[软件]。https://originality.ai

R. H. Perlis 等人（2024年）。对医生和医院的信任度下降：2020—2024年趋势。*JAMA Network Open*。

Search Engine Journal.（2025年）。*ChatGPT* 实时网络搜索能力。https://www.searchenginejournal.com

Y. Shahsavar 和 A. Choudhury（2023年）。用户对 ChatGPT 用于自我诊断和医疗决策的信心。*Digital Health*。

W. M. Silberg 等（1997年）。评估、控制并确保互联网上医疗信息的质量。*JAMA*，*277*(15)，1244-1245。https://doi.org/10.1001/jama.1997.03540390074039

Y. Sun、Y. Zhang、J. Gwizdka 和 C. B. Trace（2019年）。消费者对在线健康信息质量的评价：相关标准和指标的系统性文献综述。*Journal of Medical InternetResearch*，*21*(5)，e12522。https://doi.org/10.2196/12522

Supabase Inc.（2024年）。*Supabase*：开源 *Firebase* 替代方案 [软件]。https://supabase.com

J. Tangsrivimol 等（2025年）。医疗情境中 AI 聊天机器人的误导性信息与幻觉。*Health Informatics Journal*。

A. Tong 等（2023年）。健康错误信息与无效治疗的采用：系统综述。*Preventive Medicine*。

T. Wu 等（2025年）。大语言模型生成的健康答复与引用来源的一致性。*Nature Communications*。

表 1

AI 引用的健康信息来源的权威性信号特征（*N=615*）

权威性信号	n（%）或中位数（四分位距）
机构所属领域	
医疗机构	188 (30.6%)
政府资源	119 (19.3%)
商业健康平台	76 (12.4%)
专业/执业网站	73（11.9%）
百科全书来源	66（10.7%）
专业协会	53（8.6%）
同行评审期刊	35（5.7%）
新闻/媒体	5（0.8%）
作者资历领域	
无署名	398（64.7%）
基本署名（仅姓名）	174 (28.3%)
完整署名（姓名 + 资质）	43 (7.0%)
质量保证领域	
已注明医学审查	182（29.6%）
引用参考文献	241（39.2%）
内容时效性 - 2024—2026年（近期）	224（36.4%）
内容时效性 - 2020—2023年	141（22.9%）
内容时效性 - 2020年之前／无日期	250（40.7%）
数字权威域名	
页面权威度	中位数：55（四分位距：40-67）
域名权威度	中位数：89（四分位距：80-95）
垃圾评分	中位数：2（四分位距：1-4）
存在架构标记	457（74.3%）

内容长度——简短（<500 词）	114（18.5%）
内容长度——中等（500-1500 词）	207（33.7%）
内容长度——详尽（>1500 词）	294（47.8%）

注： 按域名整理的权威信号，优先考虑传统健康信息质量标志（机构隶属关系、作者资历、质量保证），然后再考虑数字／技术指标。页面权威度和域名权威度通过 Moz API 测量（0–100 分）。垃圾评分范围为 0–100（分数越高，= 表示垃圾信息信号越多）。IQR = 四分位距。

附录 A

ChatGPT 引用来源：组织类型分类

DA 分类阈值：专业/执业网站：DA ≤ 45 （DA 中位数 = 21）| 商业健康信息：DA > 45 （DA 中位数 = 73）

组织类型 1：医疗机构

定义：教学医院、大学附属医疗中心、医院系统、医学院和综合医疗提供直接患者护理的医疗机构，通常隶属于学术机构。

分类依据： 机构隶属关系和组织名称

DA 分数范围：41-93 | 中位数：67

注： *DA* 分数是此类别中来源的描述性特征

来源（n=46）| 引用总数：188

组织名称	DA评分	引用次数
梅奥诊所	92	61
克利夫兰诊所	88	60
约翰斯·霍普金斯医学中心	88	8
犹他大学健康中心	89	5
耶鲁医学	68	4
哈佛健康出版	93	3
斯坦福儿童健康医学中心	67	2
匹兹堡大学医学中心	83	2
芝加哥安与罗伯特·H·卢里儿童医院	64	2
斯坦福医疗保健	74	2
国家犹太健康中心	65	2
凯撒医疗集团	84	2
休斯顿卫理公会医院	68	2
健康指标与评估研究所	70	1
OncoLink	60	1
班纳医疗	64	1
Crossroads临终关怀与姑息治疗	49	1
纽约大学朗格尼健康中心	79	1
奥罗拉医疗保健	61	1
俄亥俄州立大学韦克斯纳医学中心	89	1
CARE医院	48	1
加州大学洛杉矶分校健康中心	73	1

组织名称	DA评分	引用次数
CityMD	55	1
爱荷华大学医疗保健	65	1
西奈山卫生系统	71	1
HealthPartners	63	1
加州大学圣迭戈分校健康	90	1
芝加哥大学医学中心	73	1
哥伦比亚大学	92	1
阿波罗医院	58	1
Vinmec 国际医院	49	1
儿童健康	62	1
南新罕布什尔大学	64	1
拉图尔医院	41	1
纪念斯隆凯特琳癌症中心	78	1
麻省总医院布里格姆医疗中心	69	1
圣路易斯儿童医院	61	1
纽约长老会医院	69	1
宾夕法尼亚大学牙科医学院	45	1
萨卢斯大学健康中心	42	1
德克萨斯大学健康中心	67	1
西北医学中心	70	1
内布拉斯加医学中心	62	1
特种外科医院	66	1
全美儿童医院	66	1
丹娜-法伯癌症研究所（哈佛医学院）	73	1

组织类型 2：政府资源

定义： 联邦、州和地方层级的官方政府卫生机构，包括 .gov 域名和国际政府卫生组织（例如世界卫生组织）。这些是由公共资金支持并承担政府问责制的实体。

分类依据：域名后缀（.gov）和机构隶属关系

DA 分数范围：39-95 | 中位数：89

注： *DA* 分数是描述性特征；政府来源通常因机构权威而具有较高的 *DA*。

来源（n=25）| 引用总数：119

组织名称	DA评分	引用次数
英国国家医疗服务体系	91	33
MedlinePlus	89	19
美国疾病控制与预防中心	94	16
healthdirect Australia	69	6
世界卫生组织	95	5
维多利亚州更佳健康频道	72	5
美国国家儿童健康与人类发展研究所	95	4
美国国家糖尿病、消化系统疾病和肾脏疾病研究所	95	3
北爱尔兰政府直接服务	73	3
美国国家精神卫生研究所	95	3
美国国家医学图书馆	89	3
爱尔兰卫生服务执行局	78	2
美国国家神经疾病与卒中研究所	95	2
NHS inform	67	2
美国国家耳聋及其他沟通障碍研究所	95	2
MyHealth Alberta	91	2
洛杉矶县公共卫生局	84	1
美国国家牙科与颅面研究所	95	1
美国国家关节炎、肌肉骨骼和皮肤病研究所	95	1
西弗吉尼亚州卫生与人力资源部	74	1
东中区卫生局	81	1
库雅荷加县ADAMHS委员会	39	1
弗吉尼亚州卫生部	87	1
美国环境保护署	92	1
HealthLink	62	1

组织类型 3：商业健康信息

定义： 面向消费者的健康内容平台和商业实体，其主要商业模式是向公众提供健康信息。这些营利性组织专注于健康内容创作和分发。

分类依据：机构隶属关系和组织名称；在情况模糊时，将域名权威度 > 45 作为额外分类依据

DA 分数范围：52-94 | 中位数：73

注： 在存在歧义的情况下，使用域名权威度阈值（*>45*）来区分商业平台和个人诊所网站

来源（n=28）| 引用总数：76

组织名称	DA评分	引用次数
Healthline	91	17
WebMD	94	13
Medical News Today（由 Healthline Media 所有）	92	12
Medscape	89	4
Verywell Mind	81	3
MSD Manuals	73	2
AFC Urgent Care Denver Cherry Creek	54	2
Healthgrades	73	2
Verywell Health	82	2
Epocrates	59	1
All About Vision	69	1
Vicks	60	1
Benadryl	52	1
Tylenol	60	1
默克公司 , Inc.	78	1
Cancer Therapy Advisor	56	1
李施德林	66	1
高露洁-棕榄公司	69	1
Ada	57	1
MedicineNet	87	1
Health.com	88	1
AMBOSS	54	1
优色林	55	1
HealthCentral LLC	83	1
BetterHelp	75	1
健康信息页面	94	1
GoodRx	74	1
饮食康复中心	60	1

组织类型 4：专业/执业网站

定义： 个人医疗诊所、独立执业者、私人诊所和小型医疗机构（通常有 1 至 5 名执业者）。这些是独立的医疗服务提供者，拥有自己的品牌网站，专注于吸引患者并为患者提供服务。

分类依据：机构隶属关系和组织名称；对于模糊案例，使用 DA ≤ 45 作为额外的分类指标

DA 分数范围：5-58 | 中位数：21

注：对于模糊案例，使用 *DA* 阈值（*≤45*）来区分个人执业/小型诊所与商业平台；较低的 *DA* 反映其线上影响力较小且更具本地化特征

来源（n=69）| 引用总数：73

组织名称	DA评分	引用次数
Refocus Corporate	33	4
Rock River Dental	23	2
Pete Walker，心理治疗硕士	45	1
OneSkin	43	1
结肠与消化健康专科 – Eastside 内镜中心	12	1
AmblyoPlay	25	1
150 Harley Street 耳鼻喉科	16	1
Aether Health	10	1
Open Lines	26	1
盖恩斯维尔皮肤科与皮肤外科	23	1
阿拉莫儿童眼科中心	5	1
Remagin	18	1
Rendia	39	1
韩国化妆品	13	1
Parris 牙齿矫正	17	1
G Smile	29	1
Valley 骨科研究所	16	1
我的更年期中心	40	1
Ability Central	28	1
Physio.co.uk	44	1
H&D物理治疗	16	1
Optometrists.org	58	1
HealyOS	12	1
阿瓦隆牙科集团	17	1
Spring View牙科	10	1
Beyond女性护理	8	1
弗吉尼亚足踝中心	15	1
骨科手术与运动医学中心	27	1

组织名称	DA评分	引用次数
达拉斯皮肤科联合诊所	22	1
Francis Hall 医生	11	1
Carda Health	26	1
Harbor	40	1
My Kidney Stone	23	1
Today's Vision Bulverde	12	1
马里兰眼科护理中心	18	1
WE C Hope	38	1
HMC Centre	20	1
Aliabadi 医生诊所	40	1
顶尖妇科医生	18	1
Clear Chemist	36	1
先进椎间盘置换脊柱修复中心	19	1
Strenge Spine	10	1
Voluson Club	40	1
AOTC Jax	14	1
First OC 皮肤科	22	1
Inspiring Smiles	19	1
弗农山牙科中心	21	1
Lincoln 牙科联合诊所	26	1
Ripon 牙科	7	1
Liv 牙髓病诊所	12	1
Tompkins 牙科	30	1
West Bell 牙科护理	21	1
Gloss 牙科	9	1
Texas 牙齿矫正	17	1
All In 1 牙科	25	1
iO 牙科	23	1
眼科医师	24	1
First Choice 足病诊疗	5	1
先进骨科中心：马里兰骨科专科部门	29	1
Watauga 骨科	25	1
Newtown 足踝专科	21	1

组织名称	DA评分	引用次数
Eye Doctors - Elgart Gordon & Associates	19	1
足部急诊护理	12	1
Femmy诊所	30	1
DaSC	16	1
7 Day居家护理	40	1
Spine.md	26	1
West Broward眼科协会	20	1
Piedmont儿科	18	1

组织类型5：百科全书

定义：提供全面健康信息并以百科全书格式编排的综合参考百科全书和医学百科全书。这些平台主要由用户贡献内容，或采用众包方式提供参考资料。

分类依据：机构隶属关系和百科全书格式

DA 分数范围：97-97 | 中位数：97

注：*DA* 评分是描述性特征；维基百科的高 *DA* 反映了其已确立的网络权威性 *ty*

来源 (n=1) | 引用总数：66

组织名称	DA评分	引用次数
维基百科	97	66

组织类型 6：专业协会/资源

定义： 医疗保健专业人员的会员组织、医学专业学会、由专业人士治理的患者权益倡导组织，以及疾病专项基金会。这些通常是非营利组织，致力于在特定健康领域推进医学知识、专业标准或患者权益倡导。

分类依据： 机构隶属关系和组织名称

DA 分数范围：5-91 | 中位数：65

注： *DA* 分数是描述性特征；范围较广，反映了从成熟的全国性组织（*DA* 较高）到规模较小的专科团体（*DA* 较低）的多样性

来源（n=41）| 引用总数：53

组织名称	DA评分	引用次数
美国精神病学协会	73	4
美国皮肤病学会	80	3
美国牙科协会	65	3
美国妇产科医师学会	76	3

组织名称	DA评分	引用次数
美国家庭医师学会	80	3
DermNet NZ（新西兰皮肤病学会）	68	2
睡眠基金会	75	1
哮喘 + 英国肺脏协会	68	1
美国卒中协会	67	1
美国预防虐待儿童组织	58	1
美国骨科医师学会	70	1
美国急诊医师学会	59	1
NAMI	85	1
国际 ALS/MND 协会联盟	47	1
乌得勒支整形外科	5	1
美国风湿病学会	65	1
骨骼健康与骨质疏松基金会	65	1
英国皇家儿科与儿童健康学院	64	1
BrainFacts.org	63	1
美国全国饮食失调协会	82	1
内分泌学会	73	1
NICABM	52	1
英国硬皮病与雷诺病协会	51	1
美国验光协会	58	1
美国脑基金会	54	1
美国学习障碍资源基金会	41	1
英国国家湿疹协会	59	1
英国生育网络	56	1
美国肾脏基金会	63	1
英国糖尿病协会	67	1
Trial Net	50	1
美国国家肾脏基金会	72	1
Susan G. Komen	75	1
脓毒症联盟	60	1
美国癌症协会	91	1
美国国家脑肿瘤协会	62	1
威斯康星州姑息治疗网络	40	1

组织名称	DA评分	引用次数
北美放射学会 (RSNA)、美国放射学会（ACR）、美国放射技师协会（ASRT）	66	1
英国癌症研究中心	85	1
Target ALS	43	1
脑肿瘤慈善机构	61	1

组织类型 7：同行评审期刊

定义： 发布原创研究和系统综述的学术期刊、科学出版物及同行评审研究数据库。这些来源包括单篇期刊文章，以及收录同行评审文献的聚合平台。

分类依据： 机构隶属关系和出版形式

DA 分数范围：46-95 | 中位数：84

注： *DA* 分数是　　描述性特征；较高的 *DA* 反映了学术权威性和成熟的索引系统　　*s*

来源（n=8）| 引用总数：36

组织名称	DA评分	引用次数
美国国家生物技术信息中心 PubMed	95	28
BMJ集团	76	2
NANOS	46	1
ASM期刊	75	1
ScienceDirect	93	1
临床研究杂志	77	1
Frontiers Media	92	1
爱思唯尔有限公司	93	1

组织类型 8：新闻/媒体

定义： 涵盖健康主题的传统新闻组织和媒体机构，将健康纳入更广泛的新闻报道中。这些组织以新闻业为导向，健康只是其众多报道领域之一。

分类依据： 机构隶属关系和组织名称

DA 分数范围： 78-94 | 中位数：93

注意： *DA* 分数是描述性特征；较高的 *DA* 反映了成熟的媒体权威性

来源（n=5）| 引用总数：5

组织名称	DA评分	引用次数
The Economic Times	94	1

组织名称	DA评分	引用次数
印度时报	94	1
《纽约邮报》	93	1
News-Medical.Net	78	1
Live Science	91	1

汇总统计

分析的来源总数：223 个唯一组织 | **引用总数：**615