

· 论著 · 方法学研究 ·

# 系统评价/Meta分析检索策略：从PICO到大语言模型



黄思蝶<sup>1, 2</sup>, 崔暖<sup>1, 2</sup>, 张稳<sup>3</sup>, 曾逸笛<sup>1, 2</sup>, 李琳<sup>1, 2</sup>, 李金霞<sup>1, 2</sup>, 梁昊<sup>1, 2</sup>

1. 湖南中医药大学中医诊断研究所(长沙 410208)
2. 湖南中医药大学中医诊断学湖南省重点实验室(长沙 410208)
3. 湖南中医药大学第一附属医院老年病科(长沙 410007)

**【摘要】**文献检索策略的制定是系统评价 /Meta 分析的重要步骤之一。良好的检索策略有助于研究者准确检索到相关文献，减少漏检和误检情况，从而提升研究的可靠性和有效性。目前，大语言模型与多学科领域的融合为检索策略的制定提供了新的可能。本文介绍了从 PICO 到大语言模型的多种检索策略制定方法，探讨了其应用，指出了当前检索策略存在的不足与挑战，并对未来发展方向进行了展望。

**【关键词】**系统评价；Meta 分析；检索策略；PICO；大语言模型

**【中图分类号】**R-03      **【文献标识码】**A

Search strategies for systematic review and Meta-analysis: from PICO to large language models

HUANG Sidie<sup>1,2</sup>, CUI Nuan<sup>1,2</sup>, ZHANG Wen<sup>3</sup>, ZENG Yidi<sup>1,2</sup>, LI Lin<sup>1,2</sup>, LI Jinxia<sup>1,2</sup>, LIANG Hao<sup>1,2</sup>

1. Institute of Chinese Medicine Diagnosis, Hunan University of Chinese Medicine, Changsha 410208, China

2. Hunan Provincial Key Laboratory of Traditional Chinese Medicine Diagnostics, Hunan University of Chinese Medicine, Changsha 410208, China

3. Department of Geriatrics, The First Hospital of Hunan University of Chinese Medicine, Changsha 410007, China

Correspondence: LIANG Hao, Email: lianghao@hnucm.edu.cn

**【Abstract】**The formulation of literature retrieval strategies is one of the important steps in systematic review and Meta-analysis. A good retrieval strategy can help researchers accurately retrieve relevant literature, minimize the occurrence of missed detections and false detections, and thereby enhance the reliability and effectiveness of research. Currently, the multidisciplinary integration of large language models provides new possibilities for the formulation of retrieval strategies. This article introduces the formulation methods of various retrieval strategies from PICO to large language models, discusses the application of these methods, points out the deficiencies and challenges of current retrieval strategy generation, and looks forward to the future development direction.

**【Keywords】**Systematic review; Meta-analysis; Retrieval strategy; PICO; Large language models

DOI: [10.12173/j.issn.1004-4337.202506055](https://doi.org/10.12173/j.issn.1004-4337.202506055)

基金项目：湖南省青年科技创新人才项目(2022RC1021)

通信作者：梁昊，博士，副教授，博士研究生导师，Email: lianghao@hnucm.edu.cn

系统评价，也称为系统综述，是循证实践的基础，也是为医疗决策提供科学依据的重要途径<sup>[1-2]</sup>。其目的是对特定领域的研究现状进行系统整理和深入归纳，并做出客观性总结，要求全面涵盖该研究领域且达到最小偏倚的标准<sup>[3]</sup>。检索研究领域内相关文献的广度和准确度是系统评价的重点和难点，相关文献的遗漏或错误引用会削弱结论的科学性，甚至产生误导<sup>[4]</sup>。因此，检索策略的制定作为直接影响文献检索结果的重要环节，是影响系统评价综合质量的关键因素之一<sup>[5]</sup>。

制定检索策略的核心目的在于解决如何从海量文献中精准定位所需证据的难题，并实现召回率（recall，即查全率）与准确率（precision，即查准率）的动态平衡<sup>[6]</sup>。对检索策略方法的持续探索不仅关系单个系统评价的质量，而且构成整个循证医学证据体系可靠性的基础。当前方法学标准（如 Cochrane 手册）强调检索策略必须具备可重复性、透明度和可验证性，其设计过程需要遵循 PRISMA-S 等标准并进行严格报告<sup>[7]</sup>。为系统梳理系统评价 /Meta 分析检索策略的方法演进，本文重点围绕从传统 PICO 框架<sup>[8]</sup>到大语言模型（large language models, LLMs）辅助检索的关键发展阶段展开，分析不同方法在系统评价 /Meta 分析中的适用性与局限性，尤其在 LLMs 与人工智能技术飞速发展背景下，探讨文献检索策略制定面临的机遇与挑战，以期为构建更高效和可靠的检索方法提供参考。

## 1 概念方法与临床问题构建

传统检索策略的制定多采用概念方法（conceptual approach）<sup>[9-10]</sup>。概念方法是指在信息检索过程中，通过对检索主题的分析和理解来构建临床问题，制定一个逻辑性强且覆盖面广的检索策略，以便更有效地从大量文献中筛选出与主题相关的信息。这一

过程是确保文献全面性和准确性的重要前提，具体操作步骤<sup>[11-12]</sup>：明确研究内容并构建临床问题，初步阅读文献，列出相关检索词，运用检索方法并根据检索结果对检索策略进一步调整和优化。

明确研究内容并构建临床问题是文献识别和数据提取的关键，也是确立检索策略的基础。研究人员需要确定一个具体的临床问题，以便后续能够有针对性地确定检索词和构建初步检索式。通过阅读与研究内容相关的文献，进一步了解研究领域的背景知识、常用术语及研究热点，发现潜在的检索词和概念。根据文献内容列出与研究领域密切相关的检索词。检索词应尽可能全面（包括同义词和变体），以提高召回率，确保全面检索到相关文献。通过优化检索式，运用 AND、OR 和 NOT 等布尔逻辑运算符将检索词进行合理组合，最终构成良好的检索策略。

检索词的搜集和组合需要充足的专业知识储备和缜密的逻辑思维能力。为了将复杂的临床研究问题转化为清晰、可操作的形式，循证医学推荐使用“问题构建框架”对临床问题进行分解和重组。目前已存在多种问题构建框架，框架的选择与具体研究问题的性质和研究对象有关<sup>[13-14]</sup>，见表 1。下面详细介绍 PICO、PICOS 和 SPIDER 三种常用的问题构建框架。

### 1.1 PICO

PICO 模型是循证医学中用于构建研究问题与制定检索策略的国际通用框架。该模型通过四个基本要素——P：患者 / 参与者 / 问题（patient/population/problem）、I：干预（intervention/exposure）、C：对照（control/comparison）、O：结局（outcome），帮助研究者明确问题结构并高效定位相关证据。例如，在探讨“心衰患者使用 ACE 抑制剂与 ARBs 哪种更能降低死亡和住院风险”这一问题时，可依次对应 P（心衰患者）、I（ACE 抑

表1 问题构建框架及其适用研究类型

Table 1. Question construction framework and its applicable research types

问题构建框架	含义	适用研究类型
PICO	Patient/Population/Problem、Intervention/exposure、Control/Comparison、Outcome	定量系统评价或混合性研究评价
PICOS	Patient/Population/Problem、Intervention/exposure、Control/Comparison、Outcome、Study design	定量系统评价
SPIDER	Sample、Phenomenon of interest、Design、Evaluation、Research type	定性系统评价
SPICE	Setting、Population、Intervention、Comparison、Evaluation	护理研究
PIRD	Population、Index text、Reference test、Diagnosis of interest	诊断性试验研究
CoCoPop	Condition、Context、Population	发病率和流行研究趋势
PEO	Population、Expose of interest、Outcome of response	病因和风险研究
PICo	Population、Interest of phenomena、Context	质性研究

制剂)、C(ARBs)和O(死亡和住院风险降低)。然而, PICO模型主要应用于治疗领域, 其在预后或诊断领域的研究问题方面适用性较为有限<sup>[15]</sup>。

## 1.2 PICOS

PICOS模型是PICO模型的拓展, 即在原有模型基础上加上了S—研究设计(study design), 强调在构建研究问题时明确特定研究类型, 如随机对照试验、队列研究和病例对照研究等<sup>[16-18]</sup>。这种扩展使得PICOS模型更具针对性, 能帮助研究者明确需要纳入的研究类型, 从而筛选相关性更强的文献, 更适用于干预性研究系统评价。有学者认为, 与PICO相比, PICOS可以更稳定且可重复地纳入目标文献<sup>[19]</sup>。但过度限制研究类型可能引起选择偏倚, 从而导致遗漏相关文献。

## 1.3 SPIDER

有研究表明, PICOS并不适用于所有类型的研究<sup>[20]</sup>。2012年, Cooke等根据定性和混合研究的特点提出了SPIDER模型<sup>[21]</sup>。该模型包括五个核心要素: 研究样本(Sample, S), 即研究的参与者或目标群体, 更强调参与者的主观体验或特定背景; 感兴趣现象(Phenomenon of interest, PI), 通常指研究所关注的主观经验、观点、态度或行为, 而非特定的干预措施; 研究设计(Design, D), 如访谈、焦点小组和观察等, 强调定性研究的特征; 研究的评价内容(Evaluation, E), 通常是主观数据, 如体验、观点、信念或情绪等; 研究类型(Research type, R), 主要区分定性研究、定量研究或混合方法研究。PICO和PICOS以研究的干预措施为核心, 更常用于医学或健康科学领域, 聚焦于干预措施及其效果的客观研究; SPIDER则关注研究的现象, 评价结果更倾向于主观感受, 适用于探索参与者的主观体验、观点和信念, 常用于社会科学、心理学和护理学等领域。三种模型的对比详见表2。

传统概念方法能够帮助研究者系统化地构建检索策略, 但其仍存在一定局限性。例如, 检索词的选择主要依赖于研究者的专业知识及主观判

断<sup>[22]</sup>, 缺乏客观统一的标准; 同义词和变体等可能被遗漏, 从而影响检索的全面性。此外, 筛选检索词和制定检索策略耗费研究者大量时间与精力, 尤其当研究领域复杂或文献量庞大时, 这一问题更加突出。检索策略制定的规范性和可靠性直接影响系统评价/Meta分析的质量和完成效率, 而各模块的限制条件过多可能导致文献漏检。为减少文献的遗漏, 研究者需要通过减少检索词的数量, 如去除PICOS中的S, 必要时进一步去除C甚至I, 以扩大检索范围。然而, 这种以牺牲查准率为代价的查全率提升策略, 进一步增加了后续文献筛选的工作量。有研究表明, 全系统综述平均需要约67周完成, 即使是速览性综述(rapid review)也需要6个月左右<sup>[23-24]</sup>。

## 2 客观方法

为解决概念方法客观性和有效性不足的问题, 德国医疗质量与效率研究所采用客观方法(objective approach)构建检索策略。客观方法与概念方法精确度相似, 但灵敏度更高。它不仅能够制定高质量的检索策略, 还能提高检索过程的透明度。尽管客观方法可能受限于可用文章数量, 且需要反复测试, 但其透明化和独立工作能力的优点使其成为值得探索并应用的检索策略制定方法。具体步骤: 生成原始数据集, 将原始数据集分为开发集和验证集, 使用开发集中的文献进行检索策略的开发, 验证所得到的检索策略, 标准化记录<sup>[9, 25]</sup>, 见图1。

首先, 需要生成原始数据集。原始数据集是从大量相关文献中筛选出的高质量文献, 该数据集包含与研究主题相关的各种观点和研究成果, 能够较全面反映研究领域的现状。其次, 为确保检索策略的有效性和可靠性, 原始数据集将被随机分为两部分, 其中三分之二纳入开发集, 三分之一纳入验证集。开发集主要用于检索策略的开发和优化, 验证集主要用于评估检索策略的准确性和可靠性。接着, 使用文本挖掘对开发集内的文献进行关键词频率分

表2 PICO、PICOS和SPIDER模型对比

Table 2. Comparison of PICO, PICOS and SPIDER models

项目	PICO	PICOS	SPIDER
构成	P、I、C、O	P、I、C、O、S	S、PI、D、E、R
研究对象	患特定疾病的人群	患特定疾病的人群	参与者的主观体验或特定背景
研究问题重点	干预措施及其效果	干预措施及其效果、研究设计类型	主观体验、行为和观点
适用范围	定量/定性系统评价	定量/定性系统评价(可筛选研究设计类型)	定性系统评价(社会科学、心理学和护理学等)

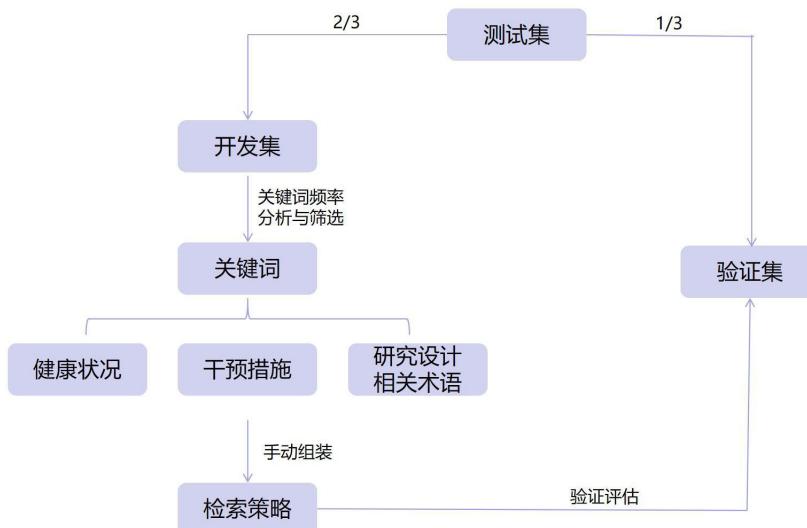


图1 客观方法步骤  
Figure 1. Steps of objective approach

析，再根据出现频率及相关度筛选关键词，分为健康状况、干预措施和研究设计相关术语三组<sup>[9]</sup>。研究者需要将这些关键词通过迭代试错法手动组装，达到敏感度与精确度的平衡统一，以避免过多不相关引用。最后，验证检索策略并标准化记录。在完成检索策略的开发后，研究者使用验证集对所得到的检索策略进行评估。评估指标包括检索的准确率和召回率等。如果检索策略在验证集上的表现达到预设标准，那么该检索策略可以被认为是有效和可靠的。为确保文献的透明度及可重复性，每个步骤都需要被标准化并详细记录，以便其他研究者能够清晰理解检索策略的制定过程。

一项系统评价纳入了 44 项关于文本挖掘在系统评价文献筛选阶段运用的研究，结果显示，文本挖掘技术可减少约 30%~70% 的人工筛选工作量<sup>[26]</sup>。然而，这种效率提升往往伴随着一定的准确性权衡，约 5% 的相关研究可能被漏检。

文本挖掘可以在检索策略制定的不同阶段进行，不仅可以自动识别相关文档中的关键术语并将其添加到生成的术语中，还可以运用“滚雪球”的方法检索文献。“滚雪球”是一种补充性检索手段，可用于发现常规搜索无法检索到的其他文献<sup>[27]</sup>。其核心在于利用已获得文献的引用关系逐步扩展检索范围，适用于检索词有限或特定领域文献离散度较高的情况。尽管该方法通常在文献检索后进行，但其结果仍可反映并优化检索策略结构与检索词选择，因此对检索策略的整体构建

具有重要的补充和验证作用。文本挖掘可通过提取、查找、获取和递归扩展四个步骤实现“滚雪球法”自动化，从而减少人工检索的时间成本并提高召回率。

此外，文本挖掘还可以进一步应用于文献筛选<sup>[28]</sup>。它能够对人工筛选的引文进行优先排序，将最相关的引文排于流程早期，提高筛选效率。部分文本挖掘程序通过融入可视化技术来进一步协助检索策略的构建，如通过热力图展示检索词与文献的相关性<sup>[29]</sup>。研究者可依据热力图颜色深浅快速识别出高频词语，并将其反馈至检索词的构建与优化中，从而提升检索质量与筛选效率。

### 3 大语言模型方法

LLMs 作为自然语言处理和机器学习领域重要研究成果，正在医学证据生成与加工全流程中逐步展现出多层次、多阶段深度嵌入的潜力<sup>[30]</sup>。随着大数据与人工智能技术的发展，LLMs 在检索策略制定中得到了全新应用。LLMs 与其他人工智能技术最显著的区别在于用户使用时不需要额外的编程指令，仅输入自然语言作为提示（prompt），LLMs 就能自动分析对话内容并生成类似于人类的回答。这大大降低了对使用者的编码能力要求，为推动多学科交叉提供了更多可能。目前，OpenAI 的 GPT-4 模型应用最为广泛。研究表明，ChatGPT-4 在摘要筛选任务中准确率可达到 90% 及以上，能以更少的人力获得可靠的结

果，是传统摘要筛选方法的替代方案之一<sup>[31]</sup>。

根据提示类型的不同，可以将其分为非引导式提示和引导式多步骤提示。非引导式提示指GPT接收到的只有单独句子或一段话指令，引导式多步骤则是指借鉴了客观方法的逻辑对GPT发出多步骤提示的指令。

### 3.1 非引导式提示

非引导式提示词可以通过单一指令完成检索策略构建或优化。根据提示词的表述不同，可以将其分为简单提示、详细提示和带示例提示三类。简单提示是仅用一句话简要说明GPT任务的提示，适用于不擅长构建复杂提示的用户。构建明确且结构化的提示词有助于GPT理解用户意图，激发大语言模型的能力，从而得到更高质量的检索策略。详细提示包含了背景信息，明确说明GPT成功完成任务所需的条件。带示例提示则是在此基础上给出一个预期的回答示例，以便GPT生成更高质量的答案。

### 3.2 引导式多步骤提示

引导式多步骤提示是基于客观方法逻辑设计的一套分步提示方案。引导式与非引导式提示均是通过收集的数据集或种子文献提取关键词，并对关键词进行分类组合以构建检索策略。不同于非引导式提示词，引导式提示将复杂的任务分解为多个简单的子任务，以便GPT理解，具体步骤见图2。研究表明，引导式多步骤提示的精度和召回率均高于非引导式提示<sup>[32]</sup>。

首先，使用一篇种子文献供GPT识别文献关键词。然后将这些关键词细分为四个不同类别：与健康状态相关（A类）、与治疗方法相关（B类）、与研究设计相关（C类），以及其他无法归入前述类别的术语（N/A类）。接着，将同类别关键词通过“OR”逻辑运算符进行组合，不同类别关键词则通过“AND”逻辑运算符联合起来，形成一个完整的检索式。为进一步提升检索的精确度和广度，可以要求GPT添加更多的关键词，如医学主题词（MeSH terms）等。引导式多步骤提示具

体示例详见表3。通过这种分步骤的引导式提示，能够更有效地控制GPT在检索构建过程中的关键词选择，并将复杂的检索任务细化为一系列更易操控且高效的子任务。与常规检索策略制定方法（如PICO等）相比，LLMs方法生成的检索式会产生大量的近义词及相关术语，从而扩大检索词范围。研究者可以通过与LLMs对话或人工进一步优化检索策略。

制定检索策略时，根据大语言模型的类型不同，提示词的构建方式也有所不同。以GPT为代表的生成式模型需依赖完整显性化的思维链（chain-of-thought, CoT），即在提示词中加入LLMs的推理步骤以提高LLMs的性能<sup>[33]</sup>。这要求研究者具备把复杂内容结构化的能力，即将Meta分析的检索要素系统整合到提示词中。前述引导式多步骤中已包含完整的CoT，能有效引导GPT分阶段输出结构化检索式。例如，在制定“脑心通胶囊治疗不稳定型心绞痛疗效的Meta分析”检索策略时，GPT的提示词见表3。

然而，这种CoT提示并不适用于以ChatGPT-o1和DeepSeek R1<sup>[34]</sup>等的推理模型。推理模型可以通过增加模型内部的CoT长度或引入强化学习（reinforcement learning, RL）机制来优化模型推理能力，从而实现模型深度思考，自主完善检索策略构建，但CoT提示会限制其推理能力的发挥。推理模型需要充分的背景信息辅助决策，因此设计层次清晰的提示词框架对任务目标的理解仍至关重要。以相同主题为例，表4列举了4种常用提示词框架并示例。

提示词框架作为模块化工具，其使用较为灵活，可以联合多个框架使用。根据系统综述检索策略生成的特点，可以将框架融合为T-R-I-A-G-E，即任务（Task）、角色（Role）、输入（Input）、行动（Action）、目标（Goal）、例子（Example）。据此框架，提示词可以优化为“你是一名经验丰富的医学信息检索专家，请你根据提供的种子文献为‘脑心通胶囊治疗不稳定型心绞痛疗效的



图2 GPT多步骤提示方案  
Figure 2. Multi-step prompt programme of GPT

表3 引导式多步骤提示示例  
Table 3. Example of a guided multi-step prompt

步骤	提示示例	大模型回复示例
第一步	请严格按照我的要求，为医学系统综述文献检索开发一个高效的检索策 略：1. 根据提供的题目及文本找出20个相关的精准检索术语。2. 请直接输 出结果不附加解释。3. 术语需满足：a)特异性强（避免泛化概念）；b)无 重复项；c)确保能有效识别新增相关研究。	1. 脑心通胶囊；2. Naoxitong Capsule；3. 不稳定性心绞痛；4. Unstable Angina；5. 急性冠状动脉综合征； 6. Acute Coronary Syndrome；7. 冠心病；8. Coronary Heart Disease；9. 冠状动脉粥样硬化性心脏病； 10. Coronary Artery Disease；11. 心绞痛发作频率；12. Angina Attack Frequency；13. 心电图缺血改变； 14. Electrocardiographic Ischemic Changes；15. 心肌缺血标志物；16. Myocardial Ischemia Biomarkers； 17. 随机对照试验；18. Randomized Controlled Trial；19. Meta分析；20. Systematic Review。
第二步	文本：“目前，冠心病已成为危害人类健康的常见疾病，我国冠心病发病率和死亡率也在呈增加趋势。不稳定型心绞痛是常见的冠心病类型，临床病情变化较快，易导致猝死或心肌梗死。在各类指南中，不稳定型心绞痛的治疗以硝酸酯类、钙通道阻滞剂及β受体阻滞剂等西药和介入治疗为主。近年来，随着许多中成药的问世，为冠心病的治疗开辟了新的途径。脑心通胶囊主要由黄芪、水蛭、地龙、全蝎、乳香、桃仁、红花、牛膝等配伍而成，具有益气活血、化瘀通络的作用，现代药理研究认为其可保护神经细胞，抑制神经细胞凋亡，保护血管内皮细胞和心肌细胞，促进侧支循环建立，改善微循环、降低心脑血管疾病发病危险因素（降血脂、降血粘）。该药自1993年上市以来，已被广泛使用，本文对其应用于不稳定型心绞痛的临床研究进行了Meta分析，以评价脑心通胶囊对不稳定型心绞痛患者的临床疗效。” 请严格遵循以下要求对第一步生成的检索术语进行分类：1. 分类框架：健 康问题（A类）、治疗干预（B类）、研究设计（C类）、其他术语（N/A类）；2. 分类标准：a)每个术语归入一个最匹配类别；b)无法归入A、B、C类的术语标注为N/A类；c)确保分类结果可复现；3. 输出格式：术语+ 分类代码（A/B/C/N/A）。请直接输出分类结果表格不附加解释。	1. 脑心通胶囊（B）；2. Naoxitong Capsule（B）；3. 不稳定性心绞痛（A）；4. Unstable Angina（A）； 5. 急性冠状动脉综合征（A）；6. Acute Coronary Syndrome（A）；7. 冠心病（A）；8. Coronary Heart Disease（A）； 9. 冠状动脉粥样硬化性心脏病（A）；10. Coronary Artery Disease（A）；11. 心绞痛发作频率（A）； 12. Angina Attack Frequency（A）；13. 心电图缺血改变（A）；14. Electrocardiographic Ischemic Changes（A）； 15. 心肌缺血标志物（A）；16. Myocardial Ischemia Biomarkers（A）；17. 随机对照试验（C）； 18. Randomized Controlled Trial（C）；19. Meta分析（C）；20. Systematic Review（C）。
第三步	请根据第二步分类结果，分别创建一个基于布尔逻辑的中文检索策略及英文检索策略。1. 逻辑结构：a)同类术语用OR连接；b)跨类组间用 AND连接；2. 所有术语需添加[Title/Abstract]后缀；3. N/A类术语不纳入检索式。请直接输出最终检索式不附加解释。例子：(itemA1[Title/ Abstract] OR itemA2[Title/ Abstract] OR itemA3[Title/ Abstract]) AND (Unstable Angina[Title/ Abstract] OR Acute Coronary Syndrome[Title/ Abstract] OR Coronary Heart Disease[Title/ Abstract] OR Coronary Artery Disease[Title/ Abstract] OR Angina Attack Frequency[Title/ Abstract] OR Electrocardiographic Ischemic Changes[Title/ Abstract] OR Myocardial Ischemia Biomarkers[Title/ Abstract]) AND (Naoxitong Capsule[Title/ Abstract] OR Randomized Controlled Trial[Title/ Abstract] OR Systematic Review[Title/ Abstract])。	中文检索式：((不稳定型心绞痛[Title/Abstract] OR 冠心病[Title/Abstract] OR 心电图缺血改变[Title/ Abstract] OR 心绞痛发作频率[Title/Abstract])) AND ((急性冠状动脉综合征[Title/Abstract] OR 冠状动脉粥样硬化性心脏病[Title/Abstract] OR 心肌缺血标志物[Title/Abstract] OR 心电图缺血改变[Title/ Abstract] OR 心绞痛[Title/Abstract] OR 心绞痛发作频率[Title/Abstract])) AND ((随机对照试验[Title/ Abstract] OR 随机对照试验[Title/Abstract] OR 系统性回顾[Title/Abstract] OR 系统性回顾[Title/ Abstract]))。

续表3

提示示例		大模型回复示例
步骤 第四步 请严格按照以下要求优化在第三步生成的检索策略：1. 结合专业知识完善 中文检索式：((不稳定型心绞痛[Title/Abstract] OR 急性冠状动脉综合征[Title/Abstract] OR 冠状动脉粥样硬化性心脏病[Title/Abstract] OR 心绞痛发作频率[Title/Abstract] OR 心电图缺血改变[Title/Abstract] OR 心肌缺血标志物[Title/Abstract] OR “冠心病” [MeSH] OR “不稳定型心绞痛” [MeSH]) AND (脑通胶囊[Title/Abstract] OR Naoxitong[Title/Abstract] OR 脑心通[Title/Abstract] OR “随机对照试验” [MeSH] OR “系统综述” [MeSH]))； 英文检索式：((Unstable Angina[Title/Abstract] OR Acute Coronary Syndrome[Title/Abstract] OR Coronary Heart Disease[Title/Abstract] OR Coronary Artery Disease[Title/Abstract] OR Angina Attack Frequency[Title/Abstract] OR Electrocardiographic Ischemic Changes[Title/Abstract] OR Myocardial Ischemia Biomarkers[Title/Abstract] OR “Angina, Unstable” [MeSH] OR “Coronary Disease” [MeSH] OR “Acute Coronary Syndrome” [MeSH] AND (Naoxitong Capsule[Title/Abstract] OR Naoxitong[Title/Abstract] OR “Naoxintong Capsule” [MeSH]) AND (Randomized Controlled Trial[Title/Abstract] OR Systematic Review[Title/Abstract] OR Meta-analysis[Title/Abstract] OR “Randomized Controlled Trials as Topic” [MeSH] OR “Systematic Reviews as Topic” [MeSH] OR “Meta-Analysis as Topic” [MeSH]))。		

表4 常用提示词框架及其示例  
Table 4. Common framework for prompt and their examples

提示词框架	具体含义	示例
R-T-F	Role（角色）、Task（任务）、Format（格式）	你是一名医学信息检索专家，专注于为系统综述开发一个高精度、高查全率的检索策略。现在你需要对“脑心通胶囊治疗不稳定型心绞痛疗效的Meta分析”进行研究，请结合PICO框架建立一个检索策略。输出格式：((A1 OR A2 OR A3) AND (B1 OR B2 OR B3) AND (C1 OR C2 OR C3))
T-A-G	Task（任务）、Action（行为）、Goal（目标）	作为系统综述研究员，你需要为“脑心通胶囊治疗不稳定型心绞痛疗效的Meta分析”开发一个检索策略。请结合PICO框架构建该检索策略。 最终生成的检索策略需满足：1. 检索策略的每个主要部分能与PICO框架对应；2. 必要时可在检索策略中添加相关的MeSH术语；3. 确保该策略能检索到所有与“脑心通胶囊治疗不稳定型心绞痛疗效的Meta分析”相关文献同时最大限度减少无关结果；4. 完整记录语扩与检索式调整过程。
C-A-R-E	Context（情景）、Action（行为）、Result（结果）、Example（例子）	你是一名经验丰富的医学信息检索专家，专注于为系统综述开发一个高精度、高查全率的检索策略。现需要为“脑心通胶囊治疗不稳定型心绞痛疗效的Meta分析”建立一个检索策略，请结合PICO框架构建该检索策略。输出格式：((A1 OR A2 OR A3) AND (B1 OR B2 OR B3) AND (C1 OR C2 OR C3))
R-I-S-E	Role（角色）、Input（输入）、Step（步骤）、Expectation（期望）	你是一名经验丰富的医学信息检索专家，请你根据提供的种子文献为“脑心通胶囊治疗不稳定型心绞痛疗效的Meta分析”建立一个检索策略。请按以下流程进行构建：运用PICO框架拆解问题、拓展术语、构建策略。确保该策略能检索到所有与“脑心通胶囊治疗不稳定型心绞痛疗效的Meta分析”相关文献同时最大限度减少无关结果（附种子文献）。

Meta 分析’建立一个检索策略。请按以下流程进行构建：运用 PICO 框架拆解问题、拓展术语、构建策略。要求确保该策略能检索到所有与‘脑心通胶囊治疗不稳定型心绞痛疗效的 Meta 分析’相关文献，同时最大限度减少无关结果。最终生成的检索策略需满足以下条件：1. 检索策略的每个主要部分能与 PICO 框架对应；2. 必要时可添加相关 MeSH 术语；3. 确保该策略能检索到所有与‘脑心通胶囊治疗不稳定型心绞痛疗效的 Meta 分析’相关文献同时最大限度减少无关结果；4. 完整记录术语扩展与检索式调整过程。例子：((A1 OR A2 OR A3) AND (B1 OR B2 OR B3) AND (C1 OR C2 OR C3))（附种子文献）”。

LLMs 方法整合了 PICO 框架与人工智能技术，既使其思考方式遵循检索策略生成的思维框架以保证科学性和完整性，又极大提升了关键词扩展和策略构建的效率。研究表明，与传统系统综述方法相比，ChatGPT 在保持准确性的同时，显著节省了时间和人力成本<sup>[35]</sup>。此外，LLMs 可以避免因疲劳导致的效率降低，具有更高的稳定性与可重复性。通过多源知识的整合，LLMs 可以满足跨学科的复杂检索需求，并通过推荐容易被忽略的近义词、同义词及相关术语，弥补人工检索时可能存在的语义盲区，从而提高检索的敏感性与文献召回率。尽管使用 GPT-4 的 API 会产生一定的费用，但是与人工相比，其效率的提升足以平衡这些成本<sup>[36]</sup>。尽管 ChatGPT 能够在一定程度上理解和生成英语之外的语言文本，但其熟练程度不一，使用英语与其交互的表现通常更可靠<sup>[37]</sup>。不同人工智能工具在语义理解、推理模式等方面存在显著差异，这导致所生成的检索策略在准确性和结果一致性上波动较大。即使是相同的提示，ChatGPT 也会生成不同的检索式，其有效性也不尽相同<sup>[38]</sup>。因此，在基于 LLMs 生成检索策略后，研究者仍需通过检索词评估、人工校验与策略优化等方法进一步提升策略的可靠性、适应性与可用性。

## 4 检索策略生成的不足

### 4.1 透明度

目前，以人工智能为主要工具的检索策略生成方法在透明度方面仍有不足，用户难以了解其内部工作机制。布尔搜索策略的复杂性进一步影

响了透明度，即使具有丰富的知识和经验，搜索者也难以制定有效的检索策略<sup>[39]</sup>。这可能导致用户对检索策略缺乏信任，进而影响使用效果。检索策略生成工具可以通过提供更多信息来获取信任，如算法原理、数据来源和参数设置等，使其更好地理解检索策略的来源和特点，提高检索效率。在提高透明度的同时也方便研究人员对生成的结果进行针对性调整。

### 4.2 可解释性

可解释性是检索策略生成的另一个关键问题。研究人员需要理解检索策略每一部分的含义和作用，以便更好地评估其效果和进行调整。然而，现有的检索策略生成工具通常忽略了可解释性，导致研究人员需要花费较多时间去理解并分析检索策略的结构。为提高可解释性，检索策略生成工具可以对检索策略进行标注加以解释和说明，例如检索策略每个部分的作用、如何影响检索结果等。此外，可以采用自然语言处理技术将检索策略翻译成易于理解的语言，从而帮助用户更好地理解。

### 4.3 可重复性

有研究表明，较多系统综述检索存在可重复性差的问题<sup>[40]</sup>。不同数据库间的检索式存在差异，检索式在应用于另一数据库时需进行术语及运算符的调整，这对检索策略的可重复性具有较大影响，研究者需要付出额外努力在不同数据库之间进行转换<sup>[11]</sup>。此外，还应考虑同一数据库在不同时间点的更新情况等因素，这可能影响检索结果完整性。为确保检索的可重复性，研究者应详细记录检索过程，包括检索的数据库、关键词、检索式、纳入标准和排除标准等全流程信息。科学利用自动化工具有助于降低人为操作可能引入的错误和偏差，从而提高研究的精确性和可信度。

## 5 总结与展望

在系统评价/Meta 分析的检索策略生成方面，目前已发展出了概念方法、客观方法和 LLMs 方法。LLMs 可以在文献筛选和信息提取等重复性任务上实现自动化，为构建高质量检索策略提供了更高效的选择。尽管 LLMs 在处理大量文献和提取关键信息方面表现出色，但其对输入的数据存在高度依赖且其受提示词长度影响，若数据不完整或提示词冗杂会影响生成结果的准确性。此

外，目前LLMs在系统评价/Meta分析制作中的运用缺乏标准化流程及伦理规范。因此，如何合理规范地运用LLMs也是目前亟待解决的问题。未来研究可以改进LLMs的训练数据和算法以提升其性能，并完善其使用规范。

需要注意的是，LLMs生成的检索方案虽能够大幅提升信息覆盖面和获取速度，但其仍依赖于已有数据与概率生成，并不具备科研人员的创新思维、批判性思维与独立验证能力。因此，过度依赖或盲目信任工具，而忽视研究者在科研过程中的核心作用，可能会导致方法上的偏差。理想的科研范式应始终以研究者为主体，人工智能为辅，人机协同。

## 参考文献

- 1 The Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions Version 5.1.0[EB/OL]. (2011-03) [2024-10-10]. <http://www.cochrane-handbook.org>
- 2 Stevens KR. Systematic reviews: the heart of evidence-based practice[J]. AACN Clin Issues, 2001, 12(4): 529–538. DOI: [10.1097/00044067-200111000-00009](https://doi.org/10.1097/00044067-200111000-00009).
- 3 Lamé G. Systematic literature reviews: an introduction[C]// Proceedings of the Design Society: International Conference on Engineering Design, 2019, 1(1): 1633–1642. DOI: [10.1017/dsi.2019.169](https://doi.org/10.1017/dsi.2019.169).
- 4 Nussbaumer-Streit B, Klerings I, Wagner G, et al. Abbreviated literature searches were viable alternatives to comprehensive searches: a meta-epidemiological study[J]. J Clin Epidemiol, 2018, 102: 1–11. DOI: [10.1016/j.jclinepi.2018.05.022](https://doi.org/10.1016/j.jclinepi.2018.05.022).
- 5 Levay P, Craven J. Systematic searching: practical ideas for improving results[M]. London: Facet Publishing, 2019: 73–94.
- 6 Sampson M, Zhang L, Morrison A, et al. An alternative to the hand searching gold standard: validating methodological search filters using relative recall[J]. BMC Med Res Methodol, 2006, 6: 33. DOI: [10.1186/1471-2288-6-33](https://doi.org/10.1186/1471-2288-6-33).
- 7 Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews[J]. Systematic reviews, 2021, 10(1): 39. DOI: [10.1186/s13643-020-01542-z](https://doi.org/10.1186/s13643-020-01542-z).
- 8 Shokraneh F. PICO framework: two decades of variation and application[EB/OL]. (2019-08-12) [2024-10-17]. <https://doi.org/10.7490/f1000research.1117334.1>
- 9 Hausner E, Waffenschmidt S, Kaiser T, et al. Routine development of objectively derived search strategies[J]. Syst Rev, 2012, 1: 19. DOI: [10.1186/2046-4053-1-19](https://doi.org/10.1186/2046-4053-1-19).
- 10 Scells H, Zuccon G, Koopman B. A comparison of automatic Boolean query formulation for systematic reviews[J]. Information Retrieval Journal, 2021, 24(1): 3–28. DOI: [10.1007/s10791-020-09381-1](https://doi.org/10.1007/s10791-020-09381-1).
- 11 Bramer WM, de Jonge GB, Rethlefsen ML, et al. A systematic approach to searching: an efficient and complete method to develop literature searches[J]. J Med Libr Assoc, 2018, 106(4): 531–541. DOI: [10.5195/jmla.2018.283](https://doi.org/10.5195/jmla.2018.283).
- 12 Carrera-Rivera A, Ochoa W, Larrinaga F, et al. How-to conduct a systematic literature review: a quick guide for computer science research[J]. MethodsX, 2022, 9: 101895. DOI: [10.1016/j.mex.2022.101895](https://doi.org/10.1016/j.mex.2022.101895).
- 13 陈耀龙, 孙雅佳, 罗旭飞, 等. 循证医学的核心方法与主要模型 [J]. 协和医学杂志 , 2023, 14(1): 1–8. [Chen YL, Sun YJ, Luo XF, et al. The core methods and key models in evidence-based medicine[J]. Medical Journal of Peking Union Medical College Hospital, 2023, 14(1): 1–8.] DOI: [10.12290/xhyzz.2022-0686](https://doi.org/10.12290/xhyzz.2022-0686).
- 14 朱政, 胡雁, 邢唯杰, 等. 不同类型循证问题的构成 [J]. 护士进修杂志 , 2017, 32(21): 1991–1994. [Zhu Z, Hu Y, Xing WJ, et al. The composition of different types of evidence based problems[J]. Journal of Nurses Training, 2017, 32(21): 1991–1994.] DOI: [10.16821/j.cnki.hsjx.2017.21.025](https://doi.org/10.16821/j.cnki.hsjx.2017.21.025).
- 15 Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions[J]. AMIA Annu Symp Proc, 2006, 2006: 359–363. <https://pubmed.ncbi.nlm.nih.gov/17238363/>
- 16 Dhrangadhariya A, Manzo G, Müller H. PICO to PICOS: weak supervision to extend datasets with new labels[J]. Stud Health Technol Inform, 2024, 316: 1775–1779. DOI: [10.3233/SHTI240775](https://doi.org/10.3233/SHTI240775).
- 17 Tacconelli E. Systematic reviews: CRD's guidance for undertaking reviews in health care[J]. The Lancet Infectious Diseases, 2010, 10(4): 226. DOI: [10.1016/S1473-3099\(10\)70065-7](https://doi.org/10.1016/S1473-3099(10)70065-7).
- 18 Methley AM, Campbell S, Chew-Graham C, et al. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews[J]. BMC Health Serv Res, 2014, 14: 579. DOI: [10.1186/s12913-014-0579-0](https://doi.org/10.1186/s12913-014-0579-0).
- 19 Saaiq M, Ashraf B. Modifying "Pico" question into "Picos" model for more robust and reproducible presentation of the methodology employed in a scientific study[J]. World J Plast Surg, 2017, 6(3): 390–392. <https://pubmed.ncbi.nlm.nih.gov/29218294/>
- 20 Anderson LM, Oliver SR, Michie S, et al. Investigating complexity in systematic reviews of interventions by using a spectrum of methods[J]. J Clin Epidemiol, 2013, 66(11): 1223–1229. DOI: [10.1016/j.jclinepi.2013.06.014](https://doi.org/10.1016/j.jclinepi.2013.06.014).
- 21 Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis[J]. Qual Health Res, 2012, 22(10): 1435–1443. DOI: [10.1177/1049732312452938](https://doi.org/10.1177/1049732312452938).
- 22 Jenkins M. Evaluation of methodological search filters--a review[J]. Health Info Libr J, 2004, 21(3): 148–163. DOI: [10.1111/j.1471-1842.2004.00511.x](https://doi.org/10.1111/j.1471-1842.2004.00511.x).
- 23 Borah R, Brown AW, Capers PL, et al. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry[J]. BMJ Open, 2017, 7(2): e012545. DOI: [10.1136/bmjopen-2017-012545](https://doi.org/10.1136/bmjopen-2017-012545).

- bmjopen-2016-012545.
- 24 Garrity C, Nussbaumer-Streit B, Hamel C, et al. Rapid reviews methods series: assessing the appropriateness of conducting a rapid review[J]. BMJ Evid Based Med, 2025, 30(1): 55–60. DOI: [10.1136/bmjebm-2023-112722](https://doi.org/10.1136/bmjebm-2023-112722).
  - 25 Hausner E, Guddat C, Hermanns T, et al. Development of search strategies for systematic reviews: validation showed the noninferiority of the objective approach[J]. J Clin Epidemiol, 2015, 68(2): 191–199. DOI: [10.1016/j.jclinepi.2014.09.016](https://doi.org/10.1016/j.jclinepi.2014.09.016).
  - 26 O'Mara-Eves A, Thomas J, McNaught J, et al. Using text mining for study identification in systematic reviews: a systematic review of current approaches[J]. Syst Rev, 2015, 4(1): 5. DOI: [10.1186/2046-4053-4-5](https://doi.org/10.1186/2046-4053-4-5).
  - 27 Choong MK, Galgani F, Dunn AG, et al. Automatic evidence retrieval for systematic reviews[J]. J Med Internet Res, 2014, 16(10): e223. DOI: [10.2196/jmir.3369](https://doi.org/10.2196/jmir.3369).
  - 28 Paynter R, Bañez LL, Berliner E, et al. EPC methods: an exploration of the use of text-mining software in systematic reviews[M]. Rockville (MD): Agency for Healthcare Research and Quality (US), 2016: 14–16. <https://pubmed.ncbi.nlm.nih.gov/27195359/>
  - 29 Mergel GD, Silveira MS, da Silva TS. A method to support search string building in systematic literature reviews through visual text mining[C]//Proceedings of the 30th Annual ACM Symposium on Applied Computing, 2015: 1594–1601. DOI: [10.1145/2695664.2695902](https://doi.org/10.1145/2695664.2695902).
  - 30 黄衍楠, 桑浩然, 刘宇, 等. 大语言模型辅助医学系统综述: 方法、发展方向和应用[J]. 广西医科大学学报, 2025, 42(3): 323–331. [Huang YN, Sang HR, Liu Y, et al. Empowering medical systematic reviews with large language models: methods, development directions, and applications[J]. Journal of Guangxi Medical University, 2025, 42(3): 323–331.] DOI: [10.16190/j.cnki.45-1211/r.2025.03.001](https://doi.org/10.16190/j.cnki.45-1211/r.2025.03.001).
  - 31 Li M, Sun J, Tan X. Evaluating the effectiveness of large language models in abstract screening: a comparative analysis[J]. Syst Rev, 2024, 13(1): 219. DOI: [10.1186/s13643-024-02609-x](https://doi.org/10.1186/s13643-024-02609-x).
  - 32 Adiga A, Wang L, Hurt B, et al. All models are useful: bayesian ensembling for robust high resolution COVID-19 forecasting[J]. medRxiv, 2021: 2021.03.12.21253495. DOI: [10.1101/2021.03.12.21253495](https://doi.org/10.1101/2021.03.12.21253495).
  - 33 Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. arXiv e-prints, 2022: arXiv:2201.11903. DOI: [10.48550/arXiv.2201.11903](https://doi.org/10.48550/arXiv.2201.11903).
  - 34 DeepSeek-AI, Guo D, Yang D, et al. DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning[J]. arXiv e-prints, 2025: arXiv:2501.12948. DOI: [10.48550/arXiv.2501.12948](https://doi.org/10.48550/arXiv.2501.12948).
  - 35 Alshami A, Elsayed M, Ali E, et al. Harnessing the power of ChatGPT for automating systematic review process: methodology, case study, limitations, and future directions[J]. Systems, 2023, 11(7): 351. <https://doi.org/10.3390/systems11070351>
  - 36 Matsui K, Utsumi T, Aoki Y, et al. Human-comparable sensitivity of large language models in identifying eligible studies through title and abstract screening: 3-layer strategy using GPT-3.5 and GPT-4 for systematic reviews[J]. J Med Internet Res, 2024, 26: e52758. DOI: [10.2196/52758](https://doi.org/10.2196/52758).
  - 37 Mahuli SA, Rai A, Mahuli AV, et al. Application ChatGPT in conducting systematic reviews and meta-analyses[J]. Br Dent J, 2023, 235(2): 90–92. DOI: [10.1038/s41415-023-6132-y](https://doi.org/10.1038/s41415-023-6132-y).
  - 38 Wang S, Scells H, Koopman B, et al. Can ChatGPT write a good boolean query for systematic review literature search?[J]. arXiv preprint, 2023, arXiv:2302.03495. DOI: [10.48550/arXiv.2302.03495](https://doi.org/10.48550/arXiv.2302.03495).
  - 39 MacFarlane A, Russell-Rose T, Shokraneh F. Search strategy formulation for systematic reviews: Issues, challenges and opportunities[J]. Intelligent Systems with Applications, 2022, 15(c): 200091. DOI: [10.1016/j.iswa.2022.200091](https://doi.org/10.1016/j.iswa.2022.200091).
  - 40 Rethlefsen ML, Brigham TJ, Price C, et al. Systematic review search strategies are poorly reported and not reproducible: a cross-sectional metaresearch study[J]. J Clin Epidemiol, 2024, 166: 111229. DOI: [10.1016/j.jclinepi.2023.111229](https://doi.org/10.1016/j.jclinepi.2023.111229).

收稿日期: 2025 年 06 月 22 日 修回日期: 2025 年 11 月 16 日

本文编辑: 张苗 黄笛

引用本文: 黄思蝶, 崔暖, 张稳, 等. 系统评价/Meta分析检索策略: 从PICO到大语言模型[J]. 数理医药学杂志, 2026, 39(1): 2–11. DOI: [10.12173/j.issn.1004-4337.202506055](https://doi.org/10.12173/j.issn.1004-4337.202506055).  
 Huang SD, Cui N, Zhang W, et al. Search strategies for systematic review and Meta-analysis: from PICO to large language models[J]. Journal of Mathematical Medicine, 2026, 39(1): 2–11. DOI: [10.12173/j.issn.1004-4337.202506055](https://doi.org/10.12173/j.issn.1004-4337.202506055).