基于轻量知识图谱与检索增强生成的药食同源候选组方设计与外部计算评价

刘卓尔[1]，孟丽苹[2#]，续晓敏[2]，李金泽[1]，李杰姝[1]，郭　杰[1]，温　烃[2*]，刘永刚[1*]

1. 北京中医药大学中药学院，北京　102400
2. 内蒙古蒙牛乳业（集团）股份有限公司，内蒙古 呼和浩特　011500

摘　要：目的　面向免疫相关药食同源候选组方，构建融合轻量知识图谱、检索增强生成（retrieval-augmented generation，RAG）和规则核验的智能辅助原型。两阶段计算评价分别用于分析方法组件的相对贡献及考察该方案在另一生成模型条件下与常用基线的差异。**方法**　知识资源包括 106 种目录物质、73 首免疫相关文献组方和 386 条来源组方正向共现关系。阶段 I 使用统一基座模型和固定参数，在 30 个经数据来源隔离核查的案例中比较无检索、语义 RAG、知识图谱辅助 RAG（GraphRAG）、量化低秩适配（quantized low-rank adaptation，QLoRA）和完整知识增强方案，并将参考组成 Jaccard 相似度（指标代码 M-P01）作为主指标。阶段 II 另行纳入 45 个比较案例，在统一通义千问（Qwen）生成条件下比较纯 Qwen、常规语义 RAG、Microsoft GraphRAG 和本研究方案。2 个阶段均进行案例层配对分析，报告 bootstrap 置信区间并作 Holm 多重比较校正；规则响应一致性则以 20 组异常输入评价。**结果**　阶段 I 的 540 次预定生成全部完成。按预设规则，评价案例与 39 份评价文本及 110 条图谱路径之间未发现案例级记录重合。B2−B1 的 M-P01 均值差为 0.078 9（95% CI: 0.045 6～0.113 6；Holm P＝0.000 369），各来源簇情景的差异方向均为正，QLoRA 则未显示独立显著增益。阶段 II 中，Q3 的 Jaccard 均值为 0.178 5；与常规语义 RAG 相比，均值差为 0.063 4（95% CI: 0.022 5～0.104 9；Holm P＝0.014 98）。其与纯 Qwen、Microsoft GraphRAG 的差异均未达统计学意义。阶段 I 规则核验改变了 2/90 次输出的组成，20 组异常输入的核验结果均与预设响应一致。**结论**　以各阶段预设的语义 RAG 为对照，阶段 I 的知识图谱辅助 RAG 和阶段 II 的本研究知识增强方案均获得更高的候选组成 Jaccard 相似度，差异具有统计学意义。这一结论仅适用于当前数据、模型和计算终点，不代表免疫功效、临床有效性或剂量适宜性，也不构成对法规合规性或安全性的一般判断。

关键词：药食同源；知识图谱；检索增强生成；候选组方设计；合规性核查；中医药信息学

中图分类号：TP18；R285　　**文献标志码：**A　　**文章编号：**0253 - 2670(2026)22 - 9035 - 13

DOI: 10.7501/j.issn.0253-2670.2026.22.018

Design and external computational evaluation of candidate formula for medicinal and edible homology substances based on lightweight knowledge graph and retrieval-augmented generation

LIU Zhuoer[1], MENG Liping[2], XU Xiaomin[2], LI Jinze[1], LI Jieshu[1], GUO Jie[1], WEN Ting[2], LIU Yonggang[1]

1. School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 102400, China
2. Inner Mongolia Mengniu Dairy (Group) Co., Ltd., Hohhot 011500, China

Abstract: Objective　To construct an intelligent auxiliary prototype integrating a lightweight knowledge graph, retrieval-augmented generation, and rule checking for candidate formulations of medicinal and edible homology related to immunity. The two-stage computational evaluation was respectively used to analyze the relative contributions of the method components and to examine the differences between this scheme and the common baseline under another generative model condition. **Methods**　Knowledge resources included 106 catalog-listed substances, 73 immune-related literature-derived formulas, and 386 co-occurrence relations derived from

收稿日期：2026-06-18

基金项目：中国食品科学技术学会食品科技专项基金（2025-M08）；国家重点研发计划（2018YFC1706500）

作者简介：刘卓尔，硕士研究生，研究方向为中药物质基础。E-mail: 18810618095@163.com

#共同第一作者：孟丽苹，博士，正高级工程师，研究方向为营养与食品卫生学。E-mail: menglipin@mengniu.cn

***通信作者：**温　烃，硕士，研究方向为产品创新与科学传播。E-mail: wenting1@mengniu.cn

刘永刚，博士，教授，研究方向为中药物质基础。E-mail: liuyg0228@163.com

source formulas. Stage I used a common base model and fixed parameters to compare no retrieval, semantic RAG, knowledge-graph-assisted RAG (GraphRAG), quantized low-rank adaptation (QLoRA), and the complete knowledge-enhanced approach in 30 cases after data-source overlap screening. Reference-composition Jaccard similarity (M-P01) was the primary endpoint. Stage II compared Qwen alone, conventional semantic RAG, Microsoft GraphRAG, and the proposed approach in 45 additional comparison cases under matched Qwen generation conditions. Both stages used paired case-level analyses, bootstrap confidence intervals, and Holm multiple comparison adjustment; 20 abnormal inputs assessed rule-response concordance. **Results** All 540 prespecified stage-I generations were completed. Under the prespecified rules, no case-level record overlap was found between the evaluation cases and the 39 evaluation documents or 110 graph paths. The mean difference of M-P01 of B2−B1 was 0.078 9 (95% CI: 0.045 6−0.113 6; Holm-adjusted P = 0.000 369), with a positive direction across source-cluster scenarios; QLoRA showed no independent significant increment. In stage II, the mean Jaccard similarity for Q3 was 0.178 5. Compared with the conventional semantic RAG, the mean difference was 0.063 4 (95% CI: 0.022 5—0.104 9; Holm-adjusted P = 0.014 98). Differences from Qwen alone and Microsoft GraphRAG were not statistically significant. In stage I, rule checking changed the output composition in 2 of 90 generations, and all 20 abnormal inputs produced the prespecified responses. **Conclusion** Taking the prespecified semantic RAG comparator for each stage as a control, candidate-composition Jaccard similarity was significantly higher for knowledge-graph-assisted RAG in stage I and the proposed knowledge-enhanced approach in stage II, and the difference was statistically significant. These findings apply only to the current data, models, and computational endpoints; they do not establish immune effects, clinical efficacy, dose appropriateness, or general judgment about regulatory compliance or safety.

Key words: medicinal and edible homology; knowledge graph; retrieval-augmented generation; candidate formula design; compliance checking; traditional Chinese medicine informatics

药食同源研究需要在传统知识利用与现代产品研究之间建立可追溯的证据联系[1]。我国相关药食同源物质范围由 2002 年原料管理通知及后续公告持续补充[2-5]，产业研究则强调资源转化、证据整合与规范化开发的重要性[6-9]。候选组方设计还需同时考虑标签、原料与剂量边界以及传统理论依据[10-13]，因此计算生成结果必须与临床处方、功效和安全性结论保持明确区分。

知识图谱为中医药实体及关系的结构化表达提供了基础[14]，检索增强生成（retrieval-augmented generation，RAG）和知识图谱辅助 RAG（GraphRAG）已被用于药食同源建议、中医药问答及知识检索[15-17]。大语言模型与图检索增强方法也逐步用于中医药知识组织和方剂生成[18-22]；相关研究进一步提出了中成药推荐、药食同源组方推荐和饮食建议生成等方法[23-27]。这些工作显示了领域知识增强的可行性，但其数据、基础模型、检索对象和评价终点并不一致。

现有研究在以下 2 方面的直接证据仍有限：一是在统一基础模型、案例和统计流程下，比较语义检索、图谱路径、量化低秩适配（quantized low-rank adaptation，QLoRA）与规则核验的相对贡献；二是在匹配生成条件下，比较领域知识增强方案与无检索大模型及常用 RAG 方案。基于此，本研究分别从组件贡献和外部基线比较 2 个层面开展评价。

本研究据此构建面向免疫相关药食同源候选组方的知识增强原型。阶段 I 设置 B0～B4 及探索性条件，估计各组件的相对贡献；阶段 II 使用另行纳入的比较案例，评价纯通义千问（Qwen）、常规语义 RAG、Microsoft GraphRAG 和本研究方案。2 阶段均以候选组成一致性为计算终点，但回答不同问题，所得效应不作合并。

1 材料与方法

1.1 研究总体设计

评价总体设计与分析流程见图 1。本候选组方原型纳入目录属性、文献组方、来源组方共现关系、文本证据和规则核验，系统组成和功能关系见图 2。评价分为 2 个阶段，分别回答上述 2 个问题。阶段 I 以受控组件比较估计语义检索、图谱路径、QLoRA 及规则核验的相对贡献；阶段 II 用另行纳入的比较案例考察知识增强方案在另一生成模型条件下相对常用基线的表现。界面和数据管理功能只用于研究实施，不作为方法有效性证据。

阶段 I 固定案例输入、基础模型、生成参数、候选域和统计流程。参考组成只在全部预定生成完成后用于评价，从未进入模型输入。阶段 II 采用另行纳入的比较案例，在相同的 Qwen 版本、生成指令、结构化输出格式和统计方案下设置 Q0～Q3。总体设计与分析流程见图 1。由于 2 个阶段的案例、基础模型和知识对象不同，结果分开报告，效应不作合并。

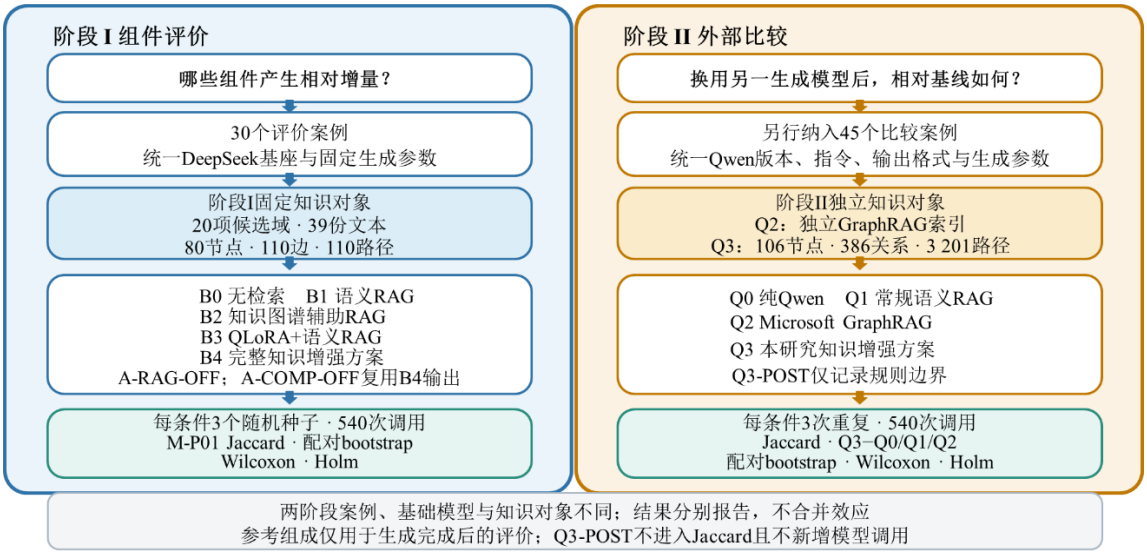


图 1　2 阶段计算评价的总体设计与分析流程

Fig. 1　Overall design and analysis flow of two-stage computational evaluation

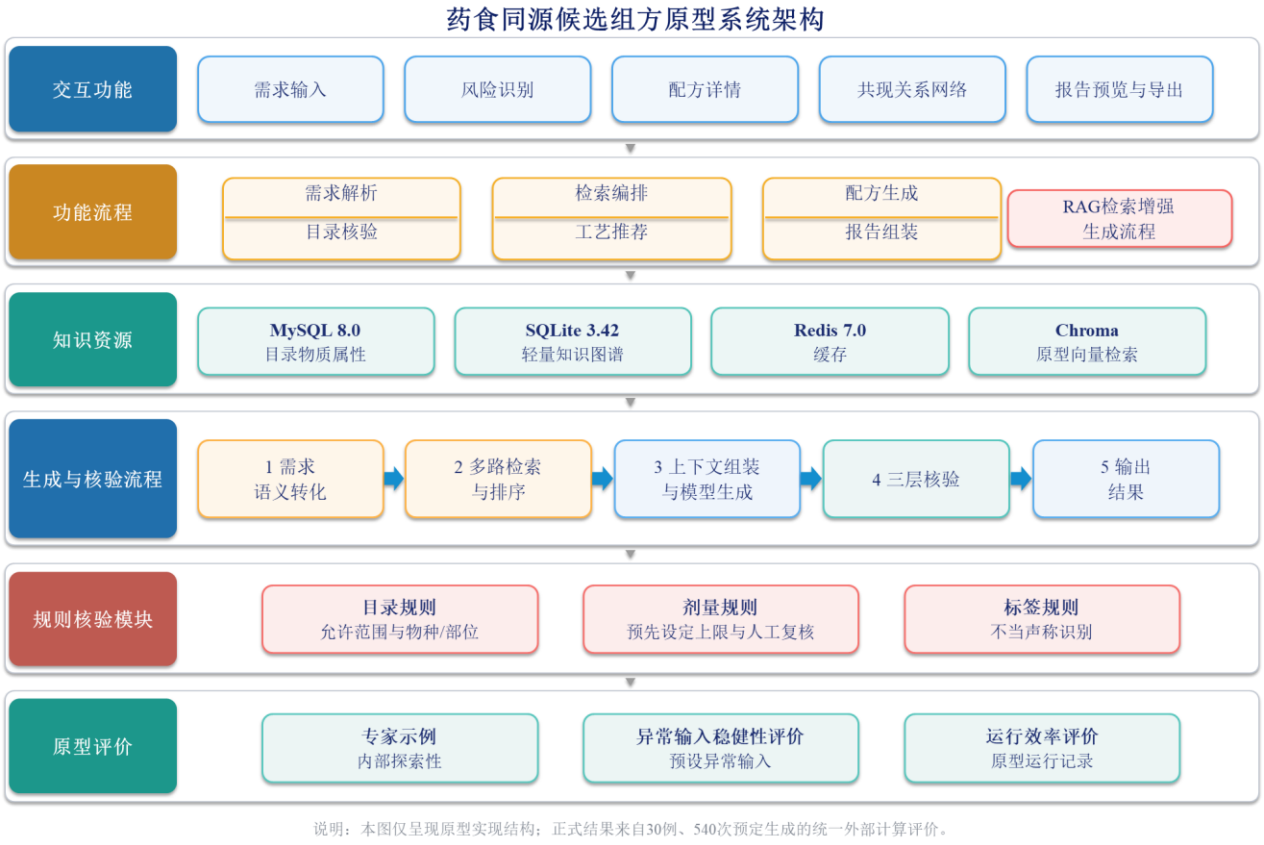


图 2　原型系统组成与功能关系

Fig. 2　Prototype system composition and functional relationships

1.2　知识资源及其在评价中的作用

1.2.1　药食同源物质与文献组方　原型知识资源以 2002 年《关于进一步规范保健食品原料管理的通知》[2]为基础，并依据 2019 年、2023 年和 2024 年公告补充的《按照传统既是食品又是中药材的物质目录》（以下简称目录）范围（106

种物质）[3-5]；参照《中国药典》2025 年版整理 106 种物质的名称和属性[11]。知识库构建阶段在中国知网和万方数据库检索免疫相关药食同源组方，检出 1 215 篇，去重后初筛纳入 183 篇，结合目录与典籍资料形成 73 首文献组方，筛选流程见图 3。双人录入一致性 Kappa 为 0.87；因缺少逐项编码表和原始计算程序，该值仅作为历史记录，不纳入主要统计分析。

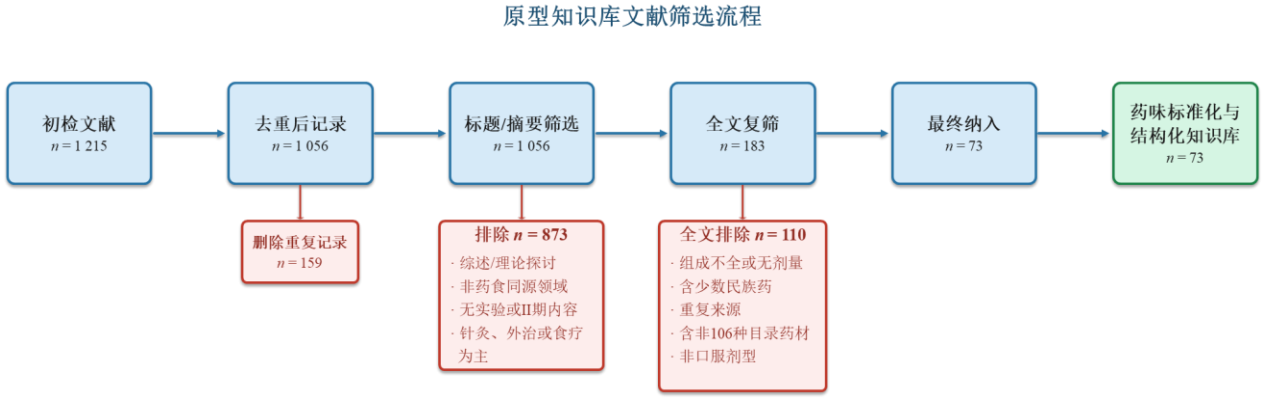


图 3　原型知识库 1 215 篇记录至 73 首文献组方的筛选流程

Fig. 3　Screening flow from prototype knowledge 1 215 records to 73 literature formulas

1.2.2　原型来源组方共现图谱构建　原型共现图谱以 106 种药食同源物质为节点，并依据来源组方中物质的共同出现整理 386 条正向共现关系。再通过检索相关文献用于研究背景和原型关系整理，不作为外部评价的疗效证据，包括相关综述为药食同源物质与研究方向提供背景[28-29]，相关实验研究涵盖复方多糖及益生菌联合组方[30-31]、4 种药食同源中药与鹿皮胶相关组方[32-33]以及复配物、炮制品和复方研究[34-36]等。图谱仅设置 1 类关系，即来源组方正向共现；该关系表示来源组方中的结构化共现，不代表药理协同、临床配伍优势或疗效证据。因此，本图谱属于任务限定的轻量领域关系图谱，而非大型通用多关系中医药知识图谱。

1.2.3　典籍与药材文本资源整理　另行整理《神农本草经》和《本草纲目》等典籍与药材说明文本[37-38]，用于原型检索和证据展示。正式评价另行固定 39 份证据文本用于语义检索；其对应证据节点及相关关系计入阶段 I 评价图谱，不计入原型共现图谱的 386 条关系。

1.2.4　阶段 I 评价知识资源　阶段 I 另行固定文本和图谱资源。评价图谱含 80 个节点、110 条边和 110 条一至二跳路径。1 跳指 2 个节点由 1 条边直接连接，2 跳指 2 个节点经 1 个中间节点和 2 条边连接。节点包括物质 20 个、证据 39 个、效应类 10 个、目录状态 5 个和风险标记 6 个；关系包括关联证据 39 条、证据对应效应 39 条、目录状态 20 条和需要复核 12 条。该图谱围绕 20 项预先限定物质及其证据、效应、目录状态和风险标记独立构建，用于阶段 I 路径检索。其在规模、关系类型和用途上均不同于含 106 个节点的原型共现图谱，也不是后者的缩减版本。原型与阶段 I 知识资源的层级、规模和用途见表 1。

1.3　阶段 I 数据角色与案例隔离

为明确训练、检索和评价数据的不同角色，73 首文献组方与 73 条 QLoRA 训练记录分别管理。依据来源记录和标准化完整组成核验，二者未见重合。阶段 I 的固定知识资源用于检索，评价案例的

表 1　原型与阶段 I 知识资源的层级、规模与评价用途

Table 1　Hierarchy, scale and evaluation purpose of prototype and phase I knowledge resources

知识资源层级	规模	结构特征	阶段与用途
原型知识层：来源组方共现图谱	106 个物质节点	386 条来源组方正向共现边；1 类关系	支持候选组方原型的来源组方关系组织；不直接用于阶段 I 效应计算
阶段 I 评价层：评价图谱	80 个节点	110 条边；4 类关系；110 条一至二跳路径	用于阶段 I 知识图谱路径检索；与原型共现图谱相互独立
阶段 I 语义检索层：评价文本	39 份文档	文本单元另行计数；对应证据节点及关系计入阶段 I 评价图谱	用于阶段 I 纯语义 RAG 及知识图谱辅助 RAG 的语义检索

参考组成在全部生成完成后才用于计算结果，各类数据按用途隔离，研究数据及其角色见表 2。

阶段 I 外部计算评价集包含 30 个案例，其中 5 例来自万方，25 例来自 Google Patents。案例与 QLoRA 训练数据在文档 ID 和标准化完整组成层面的重合数均为 0。进一步将 30 例与模型可见的 39 份评价文本及 110 条图谱路径交叉核验，文档 ID、保守归一化来源身份、配方名称和完整组成的重合数均为 0；图谱中案例级配方节点和多成分配方路径的重合数同样为 0，交叉核验结果见表 3。

表 2 研究数据与资源的角色及隔离关系

Table 2 Roles and isolation relationship of research data and resources

数据或资源	规模	形成方式	用途	与其他数据关系
原型目录物质	106 种	目录与《中国药典》资料整理	属性组织与目录规则基础	不是 QLoRA 训练集或外部评价集
原型文献组方	73 首	中国知网、万方及既有典籍整理流程	原型知识资源和共现关系构建	经来源记录和标准化完整组成核验，未发现与 QLoRA 训练记录重合
QLoRA 训练记录	73 条	训练数据为预先固定的专利披露记录	适配器训练、验证和内部测试	未使用外部评价案例；不参与 B2−B1 主要阳性比较
阶段 I 正式评价知识资源	39 份文本；80 节点、110 边、110 条一至二跳路径	正式评价前确定	语义 RAG 与知识图谱路径检索	独立建立，不等同于原型全量知识资源
阶段 I 外部计算评价案例	30 个	文档身份隔离案例集	生成与 M-P01 参考组成 Jaccard 相似度评价	案例输入用于生成；参考组成在全部生成完成后用于评价
阶段 II 外部比较案例	45 个	从冻结的公开专利案例池中另行纳入	Q0～Q3 生成与 Jaccard 比较	未参与检索资源构建；与阶段 I 正式案例未发现来源记录重合；与 Q3 资源所列专利来源标识未发现相同记录；参考组成在生成前封存

表 3 阶段 I 外部计算评价案例与正式评价知识资源的交叉核验

Table 3 Cross-validation between phase I external computational evaluation cases and formal evaluation knowledge resources

核验项	结果	解释边界
RAG 文档 ID 重合	0	预设身份核验
保守归一化来源身份重合	0	含原始精确身份，均为 0
配方名称命中	0	未检出案例级配方名称命中
完整组成命中	0	未检出参考组成完整集合命中
知识图谱案例级配方节点	0	80 节点中无外部案例配方节点
知识图谱多成分配方路径	0	110 条一/二跳路径中无案例级路径
最大组成 Jaccard	0.500 0，≥0.70 为 0，≥0.80 为 0	单物质重合不视为泄漏
数据确定顺序	评价集先行确定；知识库随后固定	不能声称知识库早于评价集确定
预先记录的场景标签	含“免疫”28/30；相邻需求 2/30	仅作描述性分类

外部参考组成与任一评价文本或图谱路径的最大组成 Jaccard 为 0.500 0，Jaccard≥0.70 或≥0.80 的对象均为 0。该最大值来自单物质层面的预期知识重合，未达到案例级配方重合的判定条件。30 例评价集先行确定，知识资源随后固定并在正式生成前保持不变。由此，本文只能确认在预设的身份、组成和路径规则下未检出案例级重合，不能排除主题或语义近邻。数据隔离核验后，依据预先记录的场景标签描述案例构成：30 例中 28 例明确包含“免疫”，另 2 例分别标记为“功能食品/药食同源组合物”和“药膳保健品/机能调节”。这些标签仅用于描述案例，不用于验证免疫功效或生物学、临床效应。

为评估案例来源相关性对主要结果的影响，开展来源簇敏感性分析。依据专利族、申请人信息和参考组成 Jaccard 相似度构建 6 种情景，组成阈值为 0.70、0.80 或 0.90，缺失申请人分别按独立或保守合并处理。各情景均采用簇内代表案例和簇均值 2 种汇总方法，并沿用 5 项预定义比较的 Wilcoxon 检验和 Holm 校正，情景 B–E 簇均值

效应量 0.049 8～0.072 6，Holm 均≤0.021 4。

1.4 阶段 I 知识图谱辅助 RAG 流程

阶段 I 采用固定离线检索流程。原型使用 Chroma 数据库和 text2vec 模型，原型递归路径查询示例见图 4；正式评价对查询和 39 份文档进行大小写归一化与分词，保留标准词项和中文单字，构造相邻中文二元词，再以稀疏词频-逆文档频率（term frequency-inverse document frequency，TF-IDF）表示并按余弦相似度排序。同分时按片段编号排序，返回前 5 名（Top 5）。图谱检索对 110 条一至二跳路径排序并返回 Top 5；GraphRAG 向模型输入 3 个语义单元和 2 条图谱路径。该流程保留原型的知识组织逻辑，但不直接复现 Chroma 向量检索，也不等同于阶段 II 使用的 Microsoft GraphRAG 实现。原型知识组织与阶段 I 正式评价方法的对应关系见表 4。

SQLite递归CTE：关系路径查询示例

```
WITH RECURSIVE compat_path(start_herb, current_herb, depth, path) AS (
    SELECT node_name, node_name, 0, node_name
    FROM kg
    WHERE node_name = '黄芪'

    UNION

    SELECT
        cp.start_herb,
        json_extract(edge.value, '$.to'),
        cp.depth + 1,
        cp.path || '→' || json_extract(edge.value, '$.to')
    FROM compat_path AS cp
    JOIN kg ON kg.node_name = cp.current_herb
    CROSS JOIN json_each(kg.edges) AS edge
    WHERE cp.depth < 3
      AND json_extract(edge.value, '$.to') IS NOT NULL
      AND json_extract(edge.value, '$.relation') != '禁忌'
)
SELECT DISTINCT current_herb FROM compat_path;
```

图 4 原型 SQLite 递归路径查询示例

Fig. 4 Example of prototype SQLite recursive path query

表 4 原型知识组织与正式评价方法的对应关系

Table 4 Correspondence between prototype knowledge organization and formal evaluation methods

项目	原型构建层	阶段 I 正式评价知识资源
目录物质	106 种属性记录	20 条预先确定的物质记录构成共享生成词表和预先记录的风险属性；并非从 30 例参考组成筛选
文献/文本	1 215 篇检索记录，经筛选形成 73 首文献组方；另整理典籍文本	从预先确定的知识资源中整理形成 39 个评价检索文档
知识图谱	106 个物质节点、386 条来源组方正向共现关系	80 个节点、110 条边、110 条一至二跳路径
语义检索	原型曾采用 text2vec 向量库	预先设定的字符二元 TF-IDF 精确余弦检索；用于确定性、离线可复现的方法组成比较，不直接复现原型 Chroma 向量检索性能
最终上下文	属性、图谱路径与文本证据组合	知识图谱辅助 RAG 取 3 个语义单元＋2 个图谱路径单元

正式评价的候选域在生成前固定为 20 条规范化物质记录，各条件共享由全部 standard_name 组成的词表。该词表不依据 30 个评价案例或其参考组成筛选，约束解码只允许输出集合内的候选名称。参考组成在 540 次生成完成后才参与指标计算。由于各条件使用同一词表，候选域不构成条件间差异来源。考虑到 20 项词表和最多输出 8 项可能降低绝对相似度，另行计算词表可达性。对案例 i，词表覆盖率为$|R_i \cap V|/|R_i|$，理论最大可达 Jaccard 为 min（$|R_i \cap V|$，8）/$|R_i|$，其中 R_i 为去重参考组成，V 为 20 项候选词表。该上限仅用于解释参考组成 Jaccard 相似度（指标代码 M-P01）绝对值和案例难度，不改变主要指标，亦不能将结果外推至 106 种原型资源或更广目录范围。

1.5 基于规则的合规性核查

模型生成后依据预先确定的规则核验目录范

围、剂量边界和标签声称。目录判定以 2002 年原料管理通知为基础[2]，并依据后续公告补充目录范围[3-5]；剂量核验参考《中国药典》[11]，标签声称核验参考 GB 7718—2011[10]。目录规则识别超出允许范围或需按物种、部位和加工方式人工复核的候选；剂量规则比较候选用量与预设上限；标签规则识别治疗、治愈等不当声称。规则类型及评价边界见表 5。自动核验不能替代具有资质人员的法规复核。

表 5 合规性核查的规则类型与评价边界

Table 5 Rule types and evaluation boundaries of compliance verification

规则类型	检查内容	当前评价边界
目录规则	候选物质是否位于预先限定范围；物种、部位和加工边界需人工复核	采用 20 组预设异常输入开展响应一致性评价；结果仅反映所设评价情形，不代表未覆盖法规场景中的总体表现
剂量规则	候选剂量是否超过预先设定上限或触发人工复核	外部案例缺少可直接比较的参考剂量及单位信息，故剂量误差不可估计
标签规则	是否含治疗、治愈等不当声称并按预设规则作出响应	20 组预设异常输入评价；不构成总体法规准确率

1.6 基座模型、量化与 QLoRA 训练

阶段 I 各条件均采用 Deepseek-AI/Deepseek-R1-Distill-Qwen-7B。QLoRA 使用 73 条预先确定的专利披露记录，按 51/11/11 划分训练、验证和内部测试集，不含 30 个外部案例；训练采用 4-bit NF4 量化，r＝8、alpha＝16、dropout＝0.1，共 3 个 epoch。检查点仅依据验证集损失选择，随后开放内部测试集；内部损失与困惑度不作为外部评价结果。

1.7 阶段 I 案例与生成参数

阶段 I 纳入 30 个案例，设置 B0～B4 和 A-RAG-OFF 共 6 个生成条件，每个条件采用 3 个随机种子（17、29、43），即 90 次模型调用，总计 540 次。A-COMP-OFF 复用 B4 的 90 次原始输出，仅关闭规则核验进行配对分析，不新增模型调用。所有预定生成均纳入分析，不按结果方向或显著性选择性重试。

1.8 阶段 I 组件评价条件

阶段 I 的主要比较条件为 B0～B4：B0 为无检索基线，B1 为纯语义 RAG，B2 为 GraphRAG，B3 为 QLoRA 联合语义 RAG，B4 为 QLoRA、GraphRAG 和规则核验组成的完整知识增强方案。A-RAG-OFF 探索关闭语义检索后的表现；A-COMP-OFF 比较同一 B4 输出在规则核验前后的组成变化。各生成条件、调用数及分析角色见表 6。2 项探索性分析均不纳入 5 项主要比较的 Holm 校正族。

表 6 阶段 I 生成条件、模型调用数与分析角色

Table 6 Generation conditions, model call counts and analysis roles of phase I

条件	LoRA	检索设置	规则核验	模型调用数	分析角色
B0	关闭	无检索	关闭	90	主要条件：无检索基线
B1	关闭	纯语义 RAG	关闭	90	主要条件：估计语义检索的相对增量
B2	关闭	GraphRAG	关闭	90	主要条件：估计图谱路径检索的相对增量
B3	开启	纯语义 RAG	关闭	90	主要条件：估计 QLoRA 的相对增量
B4	开启	GraphRAG	开启	90	主要条件：完整知识增强方案
A-RAG-OFF	开启	仅图谱路径检索	开启	90	探索性条件：关闭语义检索
A-COMP-OFF	沿用 B4	复用 B4 原始输出	关闭	0	探索性分析：规则核验前后配对比较

1.9 阶段 I 评价指标与统计分析

主指标 M-P01 为生成组成与参考组成的 Jaccard 相似度。两者均以 standard_name 标准化并集合化，重复项去除；固定别名表仅映射经审定的确切同义名称。炮制品、不同物种和不同药用部位不自动视为等同；无效候选和空并集记 0，评价时不进行案例级人工改名或调分。参考组成为统一比较基准，不代表唯一正确答案。

以案例为统计单位，先对每个案例的 3 个随机种子取均值。5 项主要比较为 B1−B0、B2−B1、B3−B1、B4−B2 和 B4−B3。采用双侧精确配对 Wilcoxon 符号秩检验：绝对差≤1×10^{-15} 的零差值不参与秩和计算，但保留在完整案例分母中并报告为相等案例；并列绝对差取平均秩。另进行 10 000

次案例层配对有放回 bootstrap，并对 5 项比较作 Holm 校正。Bootstrap 均值差区间与 Wilcoxon 秩检验评价的对象不同，显著性以预定义的 Holm 校正结果为准。统计实现经独立穷举程序及 Python 科学计算库 SciPy 的可比情形交叉验证。

为评估词表可达性对 M-P01 绝对值的影响，补充事后条件化 Jaccard 敏感性分析。生成组成保持不变，仅将参考集合限定为 $R_i \cap V$，并沿用主分析的集合化、空并集记 0、案例内 3 个随机种子聚合、Wilcoxon 检验和 10 000 次配对 bootstrap。该分析仅检验 B2−B1 在可达参考范围内的方向，不纳入主要 Holm 校正族，也不替代 M-P01。

1.10 规则响应评价与剂量边界

以 20 个预设异常输入检验规则核验的响应一致性，记录系统是否完成预期拦截或修正。所得结果仅适用于这些案例，不能用于估计未覆盖法规情形的总体准确率或产品安全性。外部案例缺少可比的参考剂量-单位对，因此剂量误差无法估计，也不报告剂量准确率、安全性或给药建议。

1.11 阶段 II 外部比较设计

阶段 II 正式案例均来源于公开专利记录，Q1 和 Q2 使用的 39 份固定文本为物质-效应证据记录。经来源标识核对，二者未发现案例级来源记录重合；阶段 II 案例筛选时，依据专利来源标识排除与 Q3 知识资源来源记录重合的案例；正式纳入的 45 例与 Q3 资源所列专利来源标识未发现相同记录。正式案例及其参考组成均未参与 Q1、Q2 或 Q3 检索资源构建，参考组成在模型生成前封存，仅在全部生成完成后用于评价。4 个条件为 Q0～Q3：Q0 为无检索的纯 Qwen[39]，Q1 为常规语义 RAG，Q2 为 Microsoft GraphRAG[40-41]，Q3 为本研究知识增强方案。每个案例在各条件下重复 3 次，共 540 次模型调用；统计单位为 45 个案例。

Qwen 版本、案例文本、生成指令、结构化输出格式、候选域、生成参数和统计方案在 4 组间保持一致，组间差别仅为知识对象和检索机制。Q0 不使用外部检索；Q1 采用常规语义 RAG；Q2 依据与 Q1 相同的 39 份固定文本独立构建 Microsoft GraphRAG 实体、关系和社区索引，并采用 Local Search；Q3 基于本研究原型知识资源形成并冻结阶段 II 专用结构化检索索引，沿用原型图谱的 106 个目录物质节点和 386 条无向来源组方正向共现边，并枚举 3 201 条一至二跳路径。对来源组方中的标准化物质集合进行两两组合，去除自连接并合并跨组方重复物质对，共得到 2 421 条候选共现物质对。目标边数按原型图谱固定为 386 条；正式模型生成前，先依据独立来源支持数降序及端点标识顺序完成节点覆盖，使 106 个节点均至少关联 1 条边，再按同一排序补足至 386 条无向共现边。“正向”仅表示观察到共同出现，不表示有向关系或药理正效应。3 201 条路径包括 386 条一跳路径和 2 815 条二跳路径；同一端点对经多个中间节点可达时，仅保留按中间节点和边标识排序最前的路径。Q3 索引来源于原型图谱；阶段 I 评价图谱和 Q2 索引则分别用于组件评价和 GraphRAG 比较，三者不混用。所得比较反映知识增强方案的整体表现，不能进一步把组间差异归因于单一检索算法、数据库规模或某一知识要素。

Q1 以 39 份固定文本为检索单元，采用 Qwen 嵌入和余弦相似度返回 Top 5；各文本按相似度降序拼接，单份文本正文最多保留 400 token，拼接后保留 1 600 token。阶段 II 的 3 次重复分别使用生成种子 501、502 和 503。4 组均采用相同的函数调用约束输出唯一 JSON（JavaScript object notation）对象，字段为 candidate_ids；候选域固定为 H001～H106，每次输出 1～8 个互不重复的稳定标识符。

阶段 II 的主要终点是生成组成与参考组成的 Jaccard 相似度。各案例先聚合 3 次重复，再完成 Q3−Q0、Q3−Q1 和 Q3−Q2 3 项预设比较。分析采用双侧配对 Wilcoxon 符号秩检验、案例层配对 bootstrap 95%置信区间（confidence interval，CI）和 Holm 校正。规则边界标记（Q3-POST）只记录在 Q3 既有输出上。该步骤不参与 Jaccard 计算，不新增模型调用，也不作为独立性能终点；生成、解析和统计方案均在正式分析前确定。

2 结果

2.1 阶段 I 评价完成度与数据隔离

阶段 I 的 6 个生成条件共完成 540 次模型调用，均得到符合预设格式的可解析输出；全部案例和随机种子均纳入分析。30 个案例与 QLoRA 训练数据在文档 ID 和标准化完整组成层面无重合，与 39 份评价文本和 110 条图谱路径交叉核验后，也未检出预设的案例级身份、名称或完整组成重合。预先记录的场景标签中，28 例明确包含“免疫”，另 2 例属于相邻候选需求。

2.2 阶段 I 组件评价结果

如表 7 和图 5 所示，纯语义 RAG 相对无检索（B1−B0）的均值差为 0.027 1（95% CI: −0.008 9～0.062 9，Holm P＝0.377），未达统计学显著。加入知识图谱路径后（B2−B1），均值差为 0.078 9（95% CI: 0.045 6～0.113 6；Holm P＝0.000 369），经 Holm 校正后达到统计学显著。相较 B1，B3 的均值差为 2.71×10^{-5}（Holm P＝1.000），未显示 QLoRA 独立增量；B4−B2 亦未达显著。B4−B3 虽为正且显著，但同时引入图谱路径和规则核验，不能分离单一组件作用。

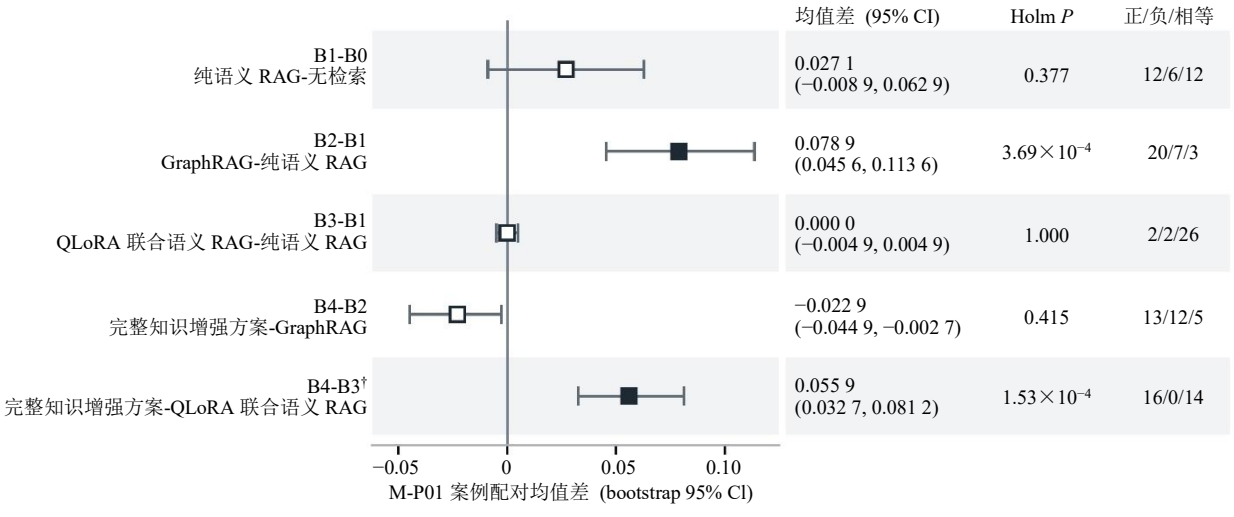
表 7 阶段 I M-P01 参考组成 Jaccard 相似度的案例配对比较

Table 7 Case pairing comparison based on Jaccard similarity of M-P01 reference composition in phase I

比较	两条件均值	均值差	未校正 95% CI	原始 P	Holm P	正/负/相等
B1−B0	0.058 5 0.031 4	0.027 100 0	−0.008 9～0.062 9	0.126 000 0	0.377 000	12/6/12
B2−B1	0.137 3 0.058 5	0.078 900 0	0.045 6～0.113 6	0.000 092 3	0.000 369	20/7/3
B3−B1	0.058 5 0.058 5	0.000 027 1	−0.004 9～0.004 9	1.000 000 0	1.000 000	2/2/26
B4−B2	0.114 4 0.137 3	−0.022 900 0	−0.044 9～−0.002 7	0.207 000 0	0.415 000	13/12/5
B4−B3†	0.114 4 0.058 5	0.055 900 0	0.032 7～0.081 2	0.000 030 5	0.000 153	16/0/14

均值按比较顺序分行；CI 为未校正配对 bootstrap 区间；显著性以 Wilcoxon-Holm P 为准；†B4−B3 同时引入知识图谱与规则核验，不能解释为单一方法组成的效应；正/负/相等为按案例配对均值差方向统计的案例数。

Means follow comparison order; CI denotes unadjusted paired bootstrap intervals; significance is determined by Wilcoxon-Holm P; †B4−B3 combines knowledge graph and rule verification, the effect cannot be attributed to a single method component ; positive/negative/equal is the number of cases counted in the direction of the mean difference of case pairing.



B0-无检索；B1-纯语义 RAG；B2-知识图谱辅助 RAG；B3-QLoRA 联合语义 RAG；B4-完整知识增强方案；方形为案例配对均值差，横线为 bootstrap 95% CI；†B4−B3 同时引入知识图谱与规则核验，不能解释为单一方法组件的效应；正/负/相等为按案例配对均值差方向统计的案例数。

B0-no retrieval; B1-pure semantic RAG; B2-knowledge graph-augmented RAG; B3-QLoRA combined semantic RAG; B4-full knowledge-enhanced scheme; squares denote the mean difference of case pairs, and horizontal lines represent the bootstrap 95% CI; †B4−B3 incorporates both knowledge graph and rule verification, and the effect cannot be attributed to a single method component; positive/negative/equal is the number of cases counted in the direction of the mean difference of case pairing.

图 5 阶段 I 不同生成策略对候选组成一致性的影响

Fig. 5 Effects of different generation strategies on consistency of candidate compositions in phase I

20 项词表对参考组成的平均覆盖率为 63.16%，最多输出 8 项时理论最大可达 Jaccard 均值为 0.629 0。在主要 Holm 校正族之外，条件化 Jaccard 的 B2−B1 均值差为 0.087 2（95% CI: 0.038 8～0.135 1；原始 P＝0.001 72），与主分析方向一致；该结果仅解释词表可达性，不替代 M-P01。

2.3　阶段 I 来源簇敏感性分析

来源簇敏感性分析中，主情景（专利族＋组成 Jaccard 0.80，缺失申请人分别保留）形成 29 个来源簇，最大簇含 2 例；纳入申请人并保守合并缺失信息时形成 22 簇，最大簇含 7 例。所有预设情景和汇总方法下 B2−B1 均为正，效应量为 0.046 7～0.072 6。最保守代表案例分析的 Holm P＝0.050 9，表明结果对极端聚类和汇总假设具有一定敏感性。情景 B–E 簇均值效应量 0.049 8～0.072 6，Holm 均≤0.021 4。

2.4　阶段 I QLoRA 适配器的相对影响

与主要比较结果一致，开启 QLoRA 的条件使用同一适配器；开启 QLoRA 并联合纯语义 RAG 的 B3 相对于未开启 QLoRA 的纯语义 RAG 条件 B1 的案例级均值差为 2.71×10^{-5}，Holm P＝1.000。在当前训练规模、任务和评价条件下，未观察到 QLoRA 对 M-P01 参考组成 Jaccard 相似度的独立显著增益。这一结果不能用于判断 QLoRA 在其他训练规模、任务或模型设置中的有效性。

2.5　规则核验与异常输入评价结果

规则核验分析复用 B4 的原始输出，并在 A-COMP-OFF 中关闭规则，不新增模型调用。规则核验使 90 次生成中的 2 次输出组成发生变化，涉及 30 个案中的 2 例：触发目录状态“物种边界需复核”后移除金钗石斛，2 项候选因此被排除；其余 299 项候选保持不变。B4−A-COMP-OFF 的案例均值差为 0.000 309（95% CI: 0～0.000 772，原始 P＝0.500，正/负/相等为 2 例/0 例/28 例）。该比较不属于主要 Holm 校正族，其结果不能外推为规则核验在其他场景中的总体表现。

20 组预设异常输入的核验结果均与预设响应一致，见表 8。15 个需拦截案例全部被识别并拦截，漏拦截为 0 例。该结果不能用于估计未测试法规场景中的总体准确率或安全性。

表 8　预设异常输入类别与评价结果

Table 8　Categories of preset anomalous inputs and evaluation results

异常输入类别	评价结果
目录边界（非目录物质、物种/部位/加工边界）	非目录物质 4/4 与预设响应一致
配伍与剂量（禁忌组合、剂量超限或需人工复核）	配伍禁忌 4/4、剂量超限 4/4 与预设一致；外部剂量误差仍不可估计
标签声称（治疗、治愈等不当声称）	标签声称 5/5 与预设响应一致
多重异常（2 类及以上同时出现）	多重异常 3/3 与预设响应一致

2.6　原型探索性辅助评价

3 名专家的内部示例、5 条固定配方评分及评分规则敏感性仅用于补充描述原型的候选组成和评分框架。由于其数据、终点和分析单位不同于 30 例主要评价，这些材料不能用于估计知识图谱、RAG 或 QLoRA 的独立影响，故不纳入主要效能推断。

其中，3 名专家的评价属于少量内部场景的非盲法示例，而 30 例外部计算评价未设置盲法多评审者量表。两者评价对象和证据等级不同，内部示例既不能替代外部评价，也不能证明外部有效性。

2.7　阶段 II 外部比较结果

阶段 II 预定的 540 次模型调用均已完成并获得符合预设结构的可解析输出，全部纳入分析；未因输出内容或结果方向进行选择性重试。在 45 个比较案例中，Q3 的 Jaccard 均值最高，为 0.178 5；Q0、Q1 和 Q2 的均值分别为 0.170 4、0.115 1 和 0.147 1。各条件结果和 3 项预设比较见表 9，完整检验信息见表 10。3 项预设比较中，Q3−Q0 的均值差为 0.008 1（95% CI: −0.034 6～0.052 9，Holm P＝0.793 6），Q3−Q1 为 0.063 4（95% CI: 0.022 5～0.104 9，Holm P＝0.014 98），Q3−Q2 为 0.031 4（95% CI: −0.011 6～0.075 3，Holm P＝0.521 1）。仅 Q3−Q1 达到统计学显著。Q3-POST 仅对既有 Q3 结果记录规则边界标记，不参与 Jaccard 计算、不新增模型调用，亦不作为独立性能终点。

3　讨论

3.1　2 阶段研究设计与总体发现

阶段 I 估计组件贡献，阶段 II 考察知识增强方案在另一生成模型下的相对表现。2 阶段研究问题、案例、基础模型和知识对象不同，结果只作互补解

表 9 阶段 II 各条件的 Jaccard 均值与预设比较

Table 9 Comparison of mean Jaccard values of each condition in phase II with preset values

类型	条件或比较	均值或均值差	95% CI	原始 *P*	Holm *P*	正/负/相等	结果解释
条件均值	Q0：纯 Qwen	0.170 4	—	—	—	—	无检索基线
条件均值	Q1：常规语义 RAG	0.1151	—	—	—	—	常规语义检索
条件均值	Q2：Microsoft GraphRAG	0.147 1	—	—	—	—	GraphRAG 检索
条件均值	Q3：本研究知识增强方案	0.178 5	—	—	—	—	领域知识与结构化检索
预设比较	Q3−Q0	0.008 1	−0.034 6～0.052 9	0.793 600	0.793 6	18/20/7	未达统计学显著
预设比较	Q3−Q1	0.063 4	0.022 5～0.104 9	0.004 993	0.015 0	27/11/7	达到统计学显著
预设比较	Q3−Q2	0.031 4	−0.011 6～0.075 3	0.260 500	0.521 1	21/19/5	未达统计学显著

各条件重复 3 次后按案例聚合；差值＝Q3－对照；95% CI 采用配对 bootstrap，*P* 采用配对 Wilcoxon 检验，Holm *P* 为多重校正后的 *P* 值；“—”表示不适用；原始 *P* 为校正前的 *P* 值；正/负/相等均按 Q3-对照的案例层均值差方向统计；零差值不参与 Wilcoxon 秩和计算，但保留在 45 例完整案例分母中。

Repeat each condition three times and then aggregate by case; difference = Q3－control; 95% CI is obtained using paired bootstrap, and *P* values are calculated using a paired Wilcoxon test and Holm *P* values are multiplicity-adjusted; “—” indicates not applicable; raw *P* values are unadjusted; positive/negative/equal counts indicate the direction of case-level mean differences (Q3 minus control); zero differences are excluded from the Wilcoxon rank calculation but retained in the denominator of 45 cases.

表 10 阶段 II 预设比较的 Wilcoxon 检验与案例差异方向

Table 10 Wilcoxon tests and case-level difference directions for phase II preset comparisons

比较	原始 P	HolmP	正/负/相等
Q3−Q0	0.793 600	0.793 60	18/20/7
Q3−Q1	0.004 993	0.014 98	27/11/7
Q3−Q2	0.260 500	0.521 10	21/19/5

释，效应不合并。

3.2 阶段 I 组件评价的解释

阶段 I 中，B1−B0 未达显著，B2−B1 为显著正向差异，B3−B1 和 B4−B2 均未达显著。B4−B3 虽为正且显著，但同时改变图谱路径和规则核验，不能归因于单一组件。因此，当前证据支持图谱路径检索在所设条件下的相对贡献，但不支持各组件均具有独立增益。

3.3 知识图谱路径检索的相对贡献

知识图谱路径检索相对纯语义检索的差异经多重校正后仍为正，各来源簇情景方向一致，提示结构化路径可能补充语义检索未充分表达的关系信息。最保守代表案例分析接近显著性边界，结果仍受来源聚类假设影响；该发现不能外推为所有知识图谱均优于语义 RAG，也不构成对 GraphRAG 的评价。

该相对增量不意味着绝对相似度较高。候选词表覆盖率和理论最大可达 Jaccard 受 20 项候选域及最多输出 8 项限制；条件化 Jaccard 虽保持正向差异，但属于事后敏感性分析。因此，阶段 I 结果仅用于解释组件间相对变化，不能判定候选组方的绝对质量。

3.4 QLoRA 与组合条件的解释边界

在当前训练规模和任务下，QLoRA 未显示独立显著增益，可能与训练记录数量及训练和外部评价粒度不同有关，不能据此外推至其他模型或任务。B4−B3 同时改变图谱路径和规则核验，A-COMP-OFF 仅观察到少量组成变化，均不足以分离单一组件作用。

3.5 免疫相关应用与规则核验边界

“免疫相关”仅界定知识资源和候选需求场景，不代表免疫功效。研究未采集细胞、动物或人体免疫终点，Jaccard 不反映生物活性，内部专家示例也不构成临床共识。规则核验仅覆盖当前输出和预设异常模式，不能外推为总体法规准确率或安全性；候选结果仍需具有资质的人员复核。

3.6 阶段 II 外部比较的解释

阶段 II 的 3 项预设比较中，只有 Q3 与 Q1 的差异达到统计学显著；Q3 与纯 Qwen、Microsoft GraphRAG 的差异均未达显著。因此，结果仅支持 Q3 在当前案例和终点下相对常规语义 RAG 的优势，不能推断其普遍优于纯大模型或 GraphRAG。4 组 Jaccard 均值为 0.115 1～0.178 5，可比较当前方法的相对表现，但不足以说明任一方案已达到较高组方质量或实际应用有效性。Q1 均值低于 Q0，提示外部检索并不必然带来一致性增益；该现象可能与语料相关性、上下文组织方式或噪声引入有

关，但本研究未分离这些因素，故不作因果解释。

阶段 I 中 B2 显著高于 B1，阶段 II 中 Q3 显著高于 Q1，均显示各自知识增强方案相对语义 RAG 的差异。两项发现仅作互补证据，效应不能合并。2 阶段的案例、基础模型、知识对象和生成方案不同，阶段 II 也不是对阶段 I 效应量的直接重复。Q3 同时涉及领域知识资源、知识组织和结构化检索，所观察到的整体差异不能进一步拆分为数据库规模、特定关系或单一检索算法的独立因果效应。

3.7 局限性与未来方向

本研究主要受评价终点和案例来源限制。参考组成是统一计算基准，却不能穷尽合理组方；Jaccard 衡量组成一致性，不能替代盲法专家评价、检索相关性评审、生物学实验或临床终点。阶段 I 虽按预设规则未发现案例级身份、名称、完整组成和路径重合，主题或语义近邻仍可能存在；阶段 II 结果也可能受案例构成与来源结构影响。

阶段 I 的 20 项候选词表限制绝对相似度，也限制结果向 106 种原型资源的外推；条件化 Jaccard 属于事后分析。阶段 II 只评价 1 个 Qwen 版本、45 个案例和固定候选域，且 Q3 与纯 Qwen、Microsoft GraphRAG 的差异未达显著，尚不能判断普遍优势。规则核验和异常输入覆盖有限，剂量误差无法估计，未采集生物学或人体证据。后续需增加来自不同来源的比较案例并完善来源分层，并开展盲法专家评价和实验验证。

4 结论

本研究构建了面向免疫相关药食同源候选组方的知识增强原型，并采用 2 阶段计算评价，分别考察组件贡献和该方案在另一生成模型条件下的相对表现。阶段 I 的知识图谱路径检索和阶段 II 的知识增强方案相较各自语义 RAG 对照，均表现出更高的候选组成 Jaccard 相似度，且差异具有统计学意义；QLoRA 未显示独立显著增益。但阶段 II 未证实本研究知识增强方案优于纯 Qwen 或 Microsoft GraphRAG。结论仅适用于当前数据、模型、候选域和计算终点，不代表免疫功效、临床有效性或剂量适宜性，也不构成对法规合规性或安全性的一般判断。

数据与代码可用性：项目研究档案保存了模型与参数、训练设置、生成策略、阶段 I 组件评价及阶段 II 外部比较资料。附录 9、10、12 列出了软件版本、生成方案、检索条件、统计方法和完整性校验，不含访问凭据或账户信息。受来源许可、第 3 方版权、专利披露和信息安全要求限制，数据、代码及模型材料不作无条件公开。在符合许可且不涉及访问凭据或账户信息的前提下，可向通信作者申请受控访问；目前尚未建立公开仓储、URL 或 DOI。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 王旭东. “药食同源”的思想源流、概念内涵与当代发展 [J]. 南京中医药大学学报, 2023, 39(9): 809-813.
[2] 卫生部. 关于进一步规范保健食品原料管理的通知(卫法监发〔2002〕51 号) [EB/OL]. (2002-03-11) [2026-08-08]. https://www.nhc.gov.cn/sps/c100088/200203/33c67ea059284ceea9e45ce61271cd12.shtml.
[3] 国家卫生健康委员会, 国家市场监督管理总局. 关于当归等 6 种新增按照传统既是食品又是中药材的物质公告 (2019 年第 8 号) [EB/OL]. (2020-01-06) [2026-08-08]. https://www.nhc.gov.cn/sps/c100088/202001/8a2f0dc747604ff9879f729bf95edf48.shtml.
[4] 国家卫生健康委员会, 国家市场监督管理总局. 关于党参等 9 种新增按照传统既是食品又是中药材的物质公告 (2023 年第 9 号) [EB/OL]. (2023-11-17) [2026-08-08]. https://www.nhc.gov.cn/sps/c100088/202311/5b062dd13fe646198b56c7d76a99aab4.shtml.
[5] 国家卫生健康委员会, 国家市场监督管理总局. 关于地黄等 4 种按照传统既是食品又是中药材的物质的公告 (2024 年第 4 号) [EB/OL]. (2024-08-26) [2026-08-08]. https://www.nhc.gov.cn/sps/c100088/202408/53150d0918ec40899b5293147ec0dd01.shtml.
[6] 康力, 葛瑞宏, 周良, 等. 我国药食同源产业发展问题解析及综合发展建议 [J]. 生命科学, 2024, 36(8): 981-990.
[7] 嵇晶, 张云羽, 王令充, 等. 基于经典名方的药食同源大健康产品开发应用探讨 [J]. 南京中医药大学学报, 2023, 39(9): 961-966.
[8] 程建明, 薛峰, 张云羽, 等. 药食同源产品研发现状、技术关键与对策 [J]. 南京中医药大学学报, 2023, 39(9): 814-826.
[9] 郭盛, 王园园, 张芳, 等. 药食同源中药资源产业现状及其开发利用策略与研究实践 [J]. 南京中医药大学学报, 2023, 39(9): 801-808.
[10] 食品安全国家标准 预包装食品标签通则: GB 7718—2011 [S].
[11] 中国药典 [S]. 一部. 2020.
[12] 中医出版中心. 黄帝内经素问 [M]. 北京: 人民卫生出版社, 2012.
[13] 李杲. 脾胃论: 新校版 [M]. 程传浩, 点校. 北京: 人民军医出版社, 2005.
[14] 贾李蓉, 刘静, 于彤, 等. 中医药知识图谱构建 [J]. 医学信息学杂志, 2015(8): 51-53.
[15] Sha H Y, Gong F, Liu B, *et al*. Leveraging retrieval-augmented large language models for dietary

recommendations with traditional Chinese medicine's medicine food homology: Algorithm development and validation [J]. *JMIR Med Inform*, 2025, 13: e75279.

[16] 董兆安, 秦可豪, 周子力, 等. 基于 GraphRAG 的中医药知识问答系统 [J/OL]. 曲阜师范大学学报 (自然科学版), 2025-04-21. https://link.cnki.net/urlid/37.1154.N.20250421.1311.003.

[17] He J L, Guo Y Q, Lam L K, *et al*. OpenTCM: A GraphRAG-empowered LLM-based system for traditional Chinese medicine knowledge retrieval and diagnosis [EB/OL]. *arXiv* [2025-04-28]. https://arxiv.org/abs/2504.20118.

[18] Shataer D, Cao S Y, Liu X, *et al*. Application of large language models in traditional Chinese medicine: A state-of-the-art review [J]. *Am J Chin Med*, 2025, 53(4): 973-997.

[19] Zhang Z X, Hao B W, Li Y J, *et al*. ZhiFangDanTai: Fine-tuning graph-based retrieval-augmented generation model for traditional Chinese medicine formula [J]. *IEEE J Biomed Health Inform*, 2026, 30(4): 3118-3131.

[20] Qu X L, Tian Z W, Cui J M, *et al*. A review of knowledge graph in traditional Chinese medicine: Analysis, construction, application and prospects [J]. *Comput Mater Continua*, 2024, 81(3): 3583-3616.

[21] 佟琳, 柴忠言, 李兵, 等. 中医药典籍知识融合推理大模型构建与应用研究 [J]. 中国中医基础医学杂志, 2025, 31(12): 2130-2135.

[22] 李敬华, 祖雅琪, 刘张怡, 等. 人工智能与大模型算法在中医脾胃病研究中的应用 [J]. 中医杂志, 2026, 67(2): 149-152.

[23] Qin S Y, Wang Y F, Cui T M, *et al*. Intelligent Chinese patent medicine (CPM) recommendation framework: Integrating large language models, retrieval-augmented generation, and the largest CPM dataset [J]. *Pharmacol Res*, 2025, 219: 107883.

[24] Li X F, He H N, Lu G B, *et al*. TCM-DS: A large language model for intelligent traditional Chinese medicine edible herbal formulas recommendations [J]. *Chin Med*, 2025, 20(1): 191.

[25] 龚凡, 刘波, 沙航宇, 等. 基于大语言模型检索增强的中医"药食同源"饮食建议生成方法 [J]. 上海中医药杂志, 2025, 59(6): 1-8.

[26] 王寒飞, 徐姝仪, 毛海洋, 等. 基于中医传承辅助平台设计用于治疗鸡滑膜炎的中药组方分析 [J]. 畜牧与兽医, 2024, 56(1): 123-128.

[27] 韩玉霞. 高血压医学医药知识图谱的构建及应用研究 [D]. 大连: 大连理工大学, 2024.

[28] Xu J Q, Zhang J L, Sang Y M, *et al*. Polysaccharides from medicine and food homology materials: A review on their extraction, purification, structure, and biological activities [J]. *Molecules*, 2022, 27(10): 3215.

[29] Hu D J, Zhang Y F, Li B Y, *et al*. Research progress on polysaccharides from medicine and food homology materials in functional foods [J]. *Chin J Nat Med*, 2025, 23(9): 1025-1035.

[30] 周钰龙, 黄晓君, 聂少平. 四君子汤复方多糖及其单方多糖的免疫调节作用 [J]. 中国食品学报, 2025, 25(5): 101-112.

[31] 刘海春, 胡晨旭, 王家明, 等. 药食同源组方联合益生菌对免疫低下小鼠的免疫增强作用 [J]. 现代食品科技, 2023, 39(2): 119-124.

[32] 高长婴, 李方琪, 严婧, 等. 淮山药等 4 种药食同源中药对免疫抑制小鼠免疫功能影响的对比研究 [J]. 中国实验诊断学, 2020, 24(5): 837-841.

[33] 靳梓微, 张荣华, 赵春娇, 等. 鹿皮胶氨基酸成分及其药食同源组方的免疫增强作用 [J]. 现代食品科技, 2025, 41(1): 37-46.

[34] 张佳俊, 廖鲜艳, 翁新楚. 灵芝黄芪复配物提高环磷酰胺对荷瘤小鼠的疗效 [J]. 上海大学学报 (自然科学版), 2012, 18(4): 425-429.

[35] 马燕, 张育贵, 石露萍, 等. 当归炮制品及其化学成分和药理作用研究进展 [J]. 中国中药杂志, 2023, 48(22): 6003-6010.

[36] 徐丽泰, 董坤, 孙银玲, 等. 基于数据挖掘与实验验证探究药食同源解酒方防治酒精性肝、脑损伤的用药规律及作用机制 [J]. 中草药, 2026, 57(12): 4742-4759.

[37] 尚志钧. 《神农本草经》校注 [M]. 北京: 学苑出版社, 2008.

[38] 李时珍. 本草纲目 (第一册) [M]. 校点本. 北京: 人民卫生出版社, 1975.

[39] Yang A, Li A F, Yang B S, *et al*. Qwen3 technical report [EB/OL]. *arXiv* [2025-05-14]. https://arxiv.org/abs/2505.09388.

[40] Edge D, Trinh H, Cheng N, *et al*. From local to global: A graph RAG approach to query-focused summarization [EB/OL]. *arXiv* [2024-04-24]. https://arxiv.org/abs/2404.16130

[41] Microsoft. GraphRAG v3.1.1 [CP/OL]. (2026-07-18) [2026-08-08]. https://github.com/microsoft/graphrag/releases/tag/v3.1.1.

[责任编辑 潘明佳]