基于AI幻觉抑制的药学智能问答平台的构建与效能验证Δ

温正旺[1*]，王嘉莹[1]，杨文月[1]，杨昊煜[1]，马　霄[2]，刘　云[1#]（1.邯郸市第一医院药学部，河北 邯郸　056003；2.东软集团有限公司，河北 邯郸　056003）

中图分类号　R95　　**文献标志码**　A　　**文章编号**　1001-0408（2026）02-0226-06
DOI　10.6039/j.issn.1001-0408.2026.02.16

摘　要　目的　构建低“人工智能（AI）幻觉”的药学智能问答平台，提升用药咨询的准确性、一致性与可追溯性。**方法**　利用Python代码对药品说明书进行批量结构化整理并构建本地药学知识库，基于大型语言模型实现检索与问答流程设计，并在Dify平台完成系统集成与本地化部署。通过设计典型临床用药问题，从达峰时间、半衰期检索及肾功能减退患者剂量调整方案推理等维度，将药学智能问答平台的输出结果与在线版DeepSeek进行对比验证，评估其检索和推理结果的准确性与可靠性。**结果**　基于本地药品说明书构建的药学智能问答平台在达峰时间、半衰期及剂量调整方案的检索和推理准确率均为100%。相比之下，在线版DeepSeek在3个维度方面的准确率分别为30%（6/20）、50%（10/20）和38%（23/60）。**结论**　构建的药学智能问答平台能够根据临床提问精准检索并提炼本地知识库信息，能避免AI幻觉的出现，为医务人员提供可靠的用药决策支持。

关键词　药学智能问答平台；AI幻觉；大型语言模型；DeepSeek；人工智能

Construction and efficacy verification of an intelligent pharmaceutical Q&A platform based on AI hallucination-suppression

WEN Zhengwang[1], WANG Jiaying[1], YANG Wenyue[1], YANG Haoyu[1], MA Xiao[2], LIU Yun[1] (1. Dept. of Pharmacy, Handan First Hospital, Hebei Handan 056003, China; 2. Neusoft Group Co., Ltd., Hebei Handan 056003, China)

ABSTRACT　OBJECTIVE To construct an intelligent pharmaceutical Q&A platform for precision medication with low “artificial intelligence (AI) hallucination”, aiming to enhance the accuracy, consistency, and traceability of medication consultations. **METHODS** Medication package inserts were batch-processed and converted into structured data through Python programming to build a local pharmaceutical knowledge base. The retrieval and question-answering processes were designed based on large language models, and system integration and localized deployment were completed on Dify platform. By designing typical clinical medication questions and comparing the output of the intelligent pharmaceutical Q&A platform with the online version of DeepSeek across dimensions such as peak time retrieval, half-life, and dosage adjustment reasoning for patients with renal impairment, the accuracy and reliability of its retrieval and reasoning results were evaluated. **RESULTS** The intelligent pharmaceutical Q&A platform, constructed based on local drug package inserts, achieved 100% accuracy in retrieval and reasoning for peak time, half-life, and dosage adjustment schemes. In comparison, the online version of DeepSeek demonstrated accuracies of 30% (6/20), 50% (10/20), and 38% (23/60) across these three dimensions, respectively. **CONCLUSIONS** The constructed intelligent pharmaceutical Q&A platform is capable of accurately retrieving and extracting information from the local knowledge base based on clinical inquiries, thereby avoiding the occurrence of AI hallucinations and providing reliable medication decision support for healthcare professionals.

KEYWORDS　intelligent pharmaceutical Q&A platform; AI hallucination; large language models; DeepSeek; artificial intelligence

近年来，随着“互联网+”与医疗大数据的快速发展，人工智能（artificial intelligence，AI）技术在医药服务领域的应用不断拓展[1]。ChatGPT、DeepSeek等大型语言模型（large language models，LLM）具备较强的自然语言处理和医学语义理解能力，在用药咨询、健康问答等任务中显示出一定潜力，可在一定程度上缓解基层医疗药学人力不足的问题。然而，通用型LLM在专业药学场景中的可靠性仍面临较大挑战。多项医学AI研究指出，LLM在处理药物相互作用、特殊人群用药、个体化剂量调整等高风险任务时，存在风险管控相对不足等情况[2]，易产生“AI幻觉”，即生成与事实不符或缺乏证据支持的内容，此类错误可能导致剂量偏差、禁忌遗漏或不当建议，对患者安全构成严重威胁。

Δ基金项目　中国青年创业就业基金会中国青年医学创新研究科研课题（No.P250320108853）

＊第一作者　副主任药师。研究方向：人工智能药学。E-mail：zhengwangwen@krae.edu.kg

#通信作者　副主任医师，硕士。研究方向：消化道恶性肿瘤、人工智能药学。E-mail：bsg_ly@hotmail.com

在此背景下，国家卫生健康委办公厅于2024年11月发布的《关于印发卫生健康行业人工智能应用场景参考指引的通知》，明确提出推进药学服务智能化的发展方向[3]，为AI在药学领域的规范化、场景化应用提供了政策依据与发展指引。政策的出台，进一步凸显了构建可靠性高、幻觉率低的药学智能问答平台的必要性。

多家医院虽已将药品说明书、诊疗指南和专家共识整合至本地数据库以供临床参考，但一些临床问题仍未得到满意答复[4-5]。究其原因，在于语言模型对已上传的本地文档语义理解不足、向量化表征效率低，导致其无法与用户查询意图进行有效匹配，从而生成了包含AI幻觉的答复。为解决上述问题，我院尝试基于LLM Transformer等算法开发一套可在基层医疗机构应用的药学智能问答平台，并利用Python代码对本地知识库进行批量标准化重构，确保LLM对本地知识库的精准检索和信息加工，从而降低AI幻觉率，提升用药咨询的准确性、一致性与可追溯性，现报道如下。

1 LLM工作原理

1.1 检索原理

在智能问答平台及本地知识库的构建过程中，先将文档进行向量化处理，再根据特定算法（如倒排索引BM25）构建检索表，存储于本地数据库（如Dify平台知识库[6]）。当用户发起查询时，LLM首先会提取问题中的关键词与语义特征，随后采用文本检索与语义检索相结合的方式完成信息匹配[7]。

1.1.1 文本检索原理

文本检索依赖于预先构建的倒排索引。在检索过程中，LLM先对候选片段（chunks——检索的最小输出单元，也是后续语义计算的基本输入单元）进行初步筛选与相关性排序，再通过余弦相似度进行精确匹配与重排。余弦相似度是一种衡量两个语句相似程度（Score值）的常用数学方法，广泛应用于自然语言处理、信息检索及机器学习领域。其核心思想是将文本转换为高维空间中的向量，通过计算两个向量夹角的余弦值来评估其相似性。Score值介于－1到1之间，越接近1，表示两个文本的语义越相似[8]。

1.1.2 语义检索原理

语义检索依赖于近似近邻检索（approximate nearest neighbor，ANN）算法，该算法将相邻的向量划分为特定区域或结构，以避免线性扫描带来的高计算成本，从而显著提升检索效率。但该机制不可避免地存在信息遗漏的风险，即部分潜在相关chunks可能因未落入查询向量划定的空间区域而未能被召回。因此，在知识库构建过程中，应尽量将语义相近的内容划分为同一chunk，而将语义差异较大的内容划分为不同chunks，以减少ANN阶段的召回失败[9]，提高知识库检索的准确性与完整性。

1.2 本地知识库向量化原理

1.2.1 向量化生成与池化过程

本地知识库上传至LLM时会经过向量化编码处理后再储存，文档在向量化前会被切分为若干chunks。chunk再被进一步分解为最小语言单元——词元（tokens）。每个token映射为初始高维向量，并通过多层Transformer的多头注意力机制依次计算其查询（Query）、键（Key）和值（Value），最终经softmax函数计算注意力权重后得到新的高维向量。

若知识库中的文本分块过长，会导致单个分块内包含多个不同主题或不相关的chunks。当将这些混合内容转换为一个整体向量时，不同方向的语义信息会在向量空间中进行“中和”或“平均”，从而稀释了核心目标语句的向量表征强度。这种“向量稀释”效应会直接导致在检索阶段关键信息的向量与用户查询向量的匹配精度下降，使得最相关的chunks难以被精准定位和召回，最终影响问答系统的准确率[10-11]，故需要通过优化分块策略来避免该问题。

1.2.2 语义完整性对向量化效果的影响

自然语言中的完整语义通常可抽象为“客体A-关系-客体B”的结构，如阿莫西林（客体A）可以抑制（关系）金黄色葡萄球菌（客体B）的繁殖。在Transformer架构中，Query更侧重于映射客体间的关系及动作，Key主要关注客体角色（即A或B），而Value则负责承载关系的具体细节（如方式、程度与结果等）。因此，在构建知识库时，确保每个chunk内“客体A-关系-客体B”语义单元的完整性，是信息能够被准确处理与检索的必要条件。同时，chunks内语句的排列顺序及上下文联系也会对检索结果产生显著影响[12]。

2 药学智能问答平台的构建

2.1 软件与硬件设备

2.1.1 软件工具

本研究使用的环境部署工具为Docker；应用开发平台为Dify；主要服务与工具包括：Rerank模型管理平台Xorbits inference，LLM DeepSeek-R1-32B、GLM4，向量化模型BGE-m3，Rerank模型BGE-Reranker-v2-m3，Python运行平台PyCharm，数据储存工具Redis、Postgres，网络请求管理工具Nginx，LLM管理平台Olalma。

2.1.2 硬件设备

本研究使用的CPU型号为I7-14700KF（美国英特尔公司），显卡为火神3090（七彩虹科技发展有限公司），硬盘为SN770 1TB M.2（美国西部数据公司），内存条为D5-6000/64G[宏基电脑（上海）有限公司]，主板为Z790-P d5（华硕电脑股份有限公司）。

2.2 软件安装和环境部署

首先安装Docker，在Docker的环境中部署Dify平台、模型管理平台Xorbits inference、网络请求管理工具Nginx和数据储存工具Redis、Postgres等；安装LLM管理平台Ollama，在Ollama上部署向量化模型BGE-m3和

LLM DeepSeek-R1-32B、GLM4。其中，Docker和Ollama的作用是创建一个应用平台的运行环境；BGE-m3用于将文本内容转化为高维向量以供LLM识别和储存，以便在本地知识库上传和用户问题解析时使用。之后，利用Python运行平台PyCharm将药品说明书等医学资料进行批量清洗，再上传至Dify平台形成本地知识库。BGE-Reranker-v2-m3把从本地知识库粗筛的结果和用户问题进行精细比对后，将Score值呈递Dify平台，在此过程中利用Nginx管理网络请求，用Redis和Postgres分别缓冲和持久储存相关数据，为后续LLM返回答案提供数据支撑。Dify平台向DeepSeek-R1-32B和GLM4发送问题后，后两者根据问题返回深度思考的答案；Dify通过应用Web供三方调用，生成应用程序，以供用户提问和向用户展示深度思考答案。软件安装和环境部署示意图见图1。

2.3 药品说明书清洗与结构设计方法

2.3.1 基于检索相似度分析的知识库结构优化

本研究在Dify平台知识库中上传了我院所有药品的说明书。为便于展示研究内容与方法，下文以伏立康唑片说明书为例进行说明，并以“伏立康唑适应证”作为测试目标。首先，依据药品说明书内容设计并构建多种不同格式的知识库结构chunks，以“伏立康唑适应证”为统一检索语句，对各结构chunks依次进行召回测试，记录各种召回情形下检索语句与召回chunks的Score值，对比分析不同知识库结构对检索质量的影响。结果见图2。

通过测试结果可以得出如下结论：药品说明书原文中，不同内容结构（如适应证、用法用量、不良反应、药理毒理等）应统一冠以药品名称，以确保语义表达的完整性；不同模块的内容需进行切分，避免混杂；在每一段的起始位置应添加概括性语句，例如“本段内容主要描述了某药的适应证/用法用量/不良反应等”，其余正文可保持不变。通过上述标准化重构，可显著提升知识库与用户问题的语义匹配精度，提高检索召回的准确性。

2.3.2 基于tokens数量控制chunks的切分优化

为验证chunks字符长度与语义相似度之间的关系，本研究随机选取3个约1 300 tokens的chunks作为实验样本。依据语义层次，自段落末尾逐步删除部分语义内容以减少字符长度，生成一系列不同长度的chunks。以统一检索语句对各长度chunks进行召回测试，并记录对应的Score值。结果（图3），当chunks长度超过约600 tokens时，匹配度出现下降趋势；而一个完整知识点的描述通常需要200～800 tokens，若切分过短则容易破坏上下文的关联，导致知识点语义不完整。综合考虑知识点完整性和召回率，本研究将药品说明书chunks的切分上限确定为500 tokens。

2.3.3 基于Python的批量清洗

根据上述药品说明书召回测试结果得到最佳药品说明书清洗结构类型。为实现对药品说明书和医学指南等知识文本的高效结构化处理，本研究引入Python语言作为主要技术工具。利用Python脚本对Excel、Word和PDF多种格式文件进行自动化处理，包括批量修改、格式转换及结构重构等。

以伏立康唑片说明书为例，由于其不同内容模块均以“【】”标注且药品名称在各模块中的复现率较低，故本研究设计药品说明书清洗方案如下：（1）提取药品说明书中【药品名称】至下一项“【”之前的文本字段为药品名称，即“x”。（2）以“x说明书”为一级标题，用“#”标记，即“#伏立康唑片说明书”；药品说明书其他部分“【】”之间的文本字段为y，以“x&y”为二级标题，用“##”标记，如“##本段内容主要描述了”&“伏立康唑片”（x）&“适应证”（y）。（3）如果二级标题下的内容超过500 tokens，则把最后一个完整语句切分至下一个chunk，避免破坏语义完整性。沿用当前二级标题+“续”作为新的标题，直至切分为不超过500 tokens的chunk。（4）将药品说明书中的表格转化成markdown格式，为确保表格完整性，表格内容不再做切分。药品说明书中表格一般位于一段完整语义表达的中间位置，为避免表格内容破坏文本上

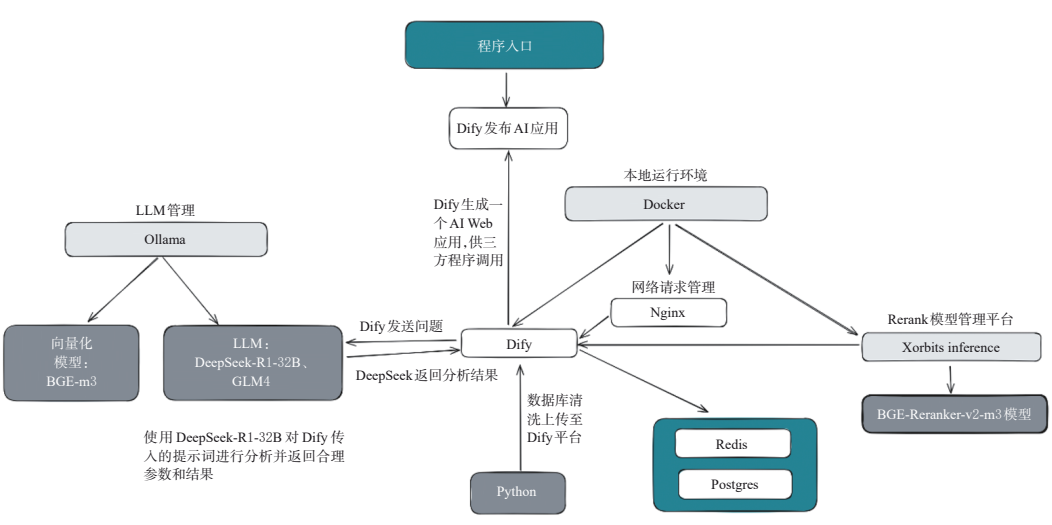


图1 软件安装和环境部署示意图

知识库结构chunks	Score值	结论
“—”+适应证+正文	召回失败	知识库chunks必须完整表示“客体A-关系-客体B”，尤其是客体的主题，否则容易召回相似的注射用伏立康唑适应证（Score值0.7），导致回答错误
伏立康唑片+“—”+正文	0.68	
伏立康唑片*1次+适应证+正文	0.73	单独客体A的频次增加会拉低Score值
伏立康唑片*5次+适应证+正文	0.71	
伏立康唑片适应证*1+正文	0.73	药品说明书语义不同的内容划分为不同的chunks，可以提高Score值
伏立康唑片适应证+正文1+伏立康唑用法用量+正文2	0.31	
伏立康唑片适应证*1+正文	0.73	“客体A+关系”组合的频次增加可以提高Score值
伏立康唑片适应证*2+正文	0.75	
伏立康唑片适应证*5+正文	0.77	
伏立康唑片适应证*5+正文	0.77	检索相关语句位于正文起始位置的Score值更高
正文1+伏立康唑片适应证... 正文5+伏立康唑适应证	0.72	
正文+伏立康唑片适应证*5	0.71	
伏立康唑片适应证*1+正文	0.73	正文起始位置添加该知识库chunks概括性语句可以提高Score值
本段内容主要描述了伏立康唑片的适应证+正文	0.75	

—：召回失败。

图2　不同知识库结构chunks的召回测试结果示例

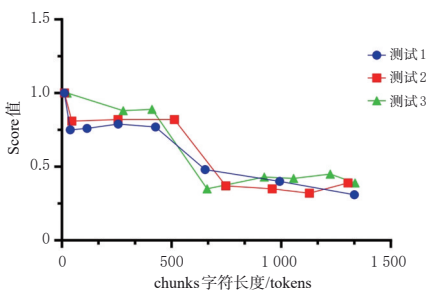


图3　不同chunks长度对应的Score值

下文的关联，应先将表格内容、表格标题、表格备注剪切处理，统一提取并编号，按照编号匹配放在对应文本的最后位置，使文本内容和表格内容分区处置。其中，表格标题增加备注信息“本段内容主要描述了”&药品名称（x）+标题，每个表格单独划分为二级标题，用“##”标记。特殊复杂表格存在一行对应多列或者一列对应多行等情况，需先行判断行列总数，确保正确的映射关系。

通过预先设定上述处理规则，我院1 120份药品说明书在2 min内完成了一键转换与标准化切分。

2.4　构建本地知识库

本研究在完成药品说明书的清洗与标准化处理后，将其以markdown格式上传至Dify平台构建本地知识库。构建过程中，分段标识符选择“##”，分段最大长度设置为4 000 tokens，并设置100 tokens的分段重叠长度以确保上下文的连续性。文本向量化采用GLM4，检索方式选择混合检索，使用BGE-Reranker-v2-m3对文本和向量检索的结果进行语义相似性匹配与重新排序，检索结果设置为Top K=10（即Score值排名前10位），Score阈值设定为0.5，以平衡召回率与准确率。

2.5　智能问答工作流的构建

在Dify平台中构建智能问答工作流，采用Chatflow模式进行流程搭建，工作流包括“开始-问题分类器-知识检索-语言模型处理-输出”。在“知识检索”节点中接入已构建的知识库；“语言模型处理”选择DeepSeek-R1-32B，并将知识检索返回的chunks作为上下文输入；经DeepSeek-R1-32B对上下文与用户问题进行语义匹配与信息提炼后，生成结构化答案并直接输出。提示词设计嵌于“语言模型处理”节点，其核心指令如下：基于所检索的知识库（即上下文）找出与问题相关的语句进行提炼，去除无关信息；如果没有找到相关信息，则反馈“本地知识库没有找到相关资料，请联系静配中心药师进行维护”，避免LLM脱离本地知识库给出不切合实际的答案。

3　平台运行情况和验证

3.1　药学智能问答平台运行及维护

我院基于Dify平台构建的药学智能问答平台实现了对本地知识库的精准检索与自动化问答。该平台能够快速定位药品说明书中的相关内容，并对检索结果进行提炼、推理和整合，生成具有临床参考价值的回答；同时，该平台可同步展示药品说明书原文，便于医务人员查阅和比对，从而保证答案的可靠性，有效降低了可能出现的“AI幻觉”风险；此外，该平台还可生成访问链接并嵌入医院信息系统的下拉菜单，实现全院范围内的实时访问与调用。

为确保平台内容的安全性与可控性，本地知识库不具备自主学习能力，并部署于医院内部网络环境，彻底隔绝外网访问带来的潜在风险。当系统识别到知识缺失时，会自动提示后台专业人员进行审核与补充，以保证知识内容的准确性、及时性与可追溯性。针对新增药品，我院实行优先更新机制，确保其药品说明书及相关资料在进院24 h内完成录入并纳入知识库，以满足临床即时查询需求。

3.2　药学智能问答平台定量检索准确性验证

药物达峰时间（t_{max}）是反映药物吸收速度的重要药动学参数，直接影响起效时间和给药方案的设计；药物半衰期（$t_{1/2}$）则用于指导给药间隔，以维持血药浓度在治疗窗内，兼顾疗效与安全性。本研究选择t_{max}与$t_{1/2}$作为客观指标，用于评价我院构建的药学智能问答平台信息检索的准确性和临床适用性。本研究从我院275种非复方口服制剂中，采用随机函数抽取20种药品，分别查阅其说明书中的t_{max}及$t_{1/2}$，并将相同问题分别输入药学智能问答平台和在线版DeepSeek进行对比。结果（表1）显示，我院药学智能问答平台给出的答案与药品说明书完全一致，t_{max}与$t_{1/2}$准确率均为100%；而在线版DeepSeek结果中t_{max}准确率为30%（6/20），$t_{1/2}$准确率为50%（10/20）。二者差异具有统计学意义（t_{max}准确率比较，χ^2=21.53，P<0.001；$t_{1/2}$准确率比较，χ^2=13.33，P<0.001），提示该平台在药学参数的检索和回答方面具有较高的准确性与可靠性。

表1　我院药学智能问答平台与在线DeepSeek在20个口服制剂t_{max}与$t_{1/2}$回答上的准确性比较

药品	说明书参数		药学智能问答平台				在线版DeepSeek			
	t_{max}/h	$t_{1/2}$/h	t_{max}/h	t_{max}与说明书是否一致	$t_{1/2}$/h	$t_{1/2}$与说明书是否一致	t_{max}/h	t_{max}与说明书是否一致	$t_{1/2}$/h	$t_{1/2}$与说明书是否一致
阿奇霉素干混悬剂	2.2	68	2.2	是	68	是	2～3	否	68	是
盐酸二甲双胍片	2.5	6.5	2.5	是	6.5	是	2～3	否	4～6	否
盐酸特拉唑嗪片	1	12	1	是	12	是	1～2	否	9～12	否
地屈孕酮片	0.5	5～7	0.5	是	5～7	是	0.5～2.5	否	5～7	是
瑞舒伐他汀钙片2	5	19	5	是	19	是	3～5	否	19	是
甲钴胺片	3.6±0.5	12.5	3.6±0.5	是	12.5	是	1～3	否	12～24	否
醋酸甲羟孕酮片	2.58±0.5	14.0±2.12	2.58±0.5	是	14.0±2.12	是	2～4	否	12～24	否
苯妥英钠片	4～12	7～42	4～12	是	7～42	是	4～12	是	7～42	是
依巴斯汀片	2.6～4	15～19	2.6～4	是	15～19	是	2～4	否	10～16	否
盐酸多奈哌齐分散片	3～4	70	3～4	是	70	是	3～4	是	70	是
孟鲁司特钠咀嚼片	2	2.7～5.5	2	是	2.7～5.5	是	2～4	否	2.5～5.5	否
左甲状腺素钠片	5～6	168	5～6	是	168	是	2～4	否	168	是
骨化三醇软胶囊	2～6	5～8	2～6	是	5～8	是	3～6	否	5～8	是
依帕司他片	1	1.8	1	是	1.8	是	1～3	否	1.5～2	否
盐酸坦索罗辛缓释胶囊	6.8	10	6.8	是	10	是	6.8	是	9～13	否
非那雄胺片	1～2	5～6	1～2	是	5～6	是	1～2	是	5～6	是
伊曲康唑胶囊	4.6±1.3	23.8±4.7	4.6±1.3	是	23.8±4.7	是	2～5	否	16～28	否
维胺酯胶囊	2.612±0.778	2.378±0.871	2.612±0.778	是	2.378±0.871	是	2～4	否	10～20	否
苯甲酸阿格列汀片	1～2	21	1～2	是	21	是	1～2	是	21	是
厄贝沙坦分散片	1.5～2	11～15	1.5～2	是	11～15	是	1.5～2	是	11～15	是
回答准确例数/总例数				20/20		20/20		6/20		10/20

3.3　药学智能问答平台定量推理准确性验证

患者肾功能减退可显著影响多种药物的药动学和药效学，若剂量调整不当，将增加不良反应发生风险。本研究选取12种需根据肌酐清除率调整剂量的药品，分别以肌酐清除率56、36、16、6 mL/min及透析状态为条件，检索我院药学智能问答平台的推荐剂量和给药频次，并与药品说明书及在线版DeepSeek的回答进行对比。结果显示，药学智能问答平台的输出答案与药品说明书一致，而在线版DeepSeek结果的准确率为38%（23/60），二者差异具有统计学意义（$P<0.001$），提示该平台在药物个体化给药推理方面具有较高准确性（限于篇幅，具体验证结果可扫描本文首页二维码进入“增强出版”板块查看附表1）。

此外，本研究还针对药物相互作用、禁忌证及特殊人群（肝功能损伤、孕妇、儿童）用药等高频临床场景进行了进一步验证，均获得准确且可用的回答（具体验证过程及结果可扫描本文首页二维码进入“增强出版”板块查看附图1～附图7），说明该药学智能问答平台已能够较好满足基层医疗机构临床药学服务需求。

4　讨论

4.1　知识库构建原理对平台性能的影响

药品说明书属于高度结构化的专业文本，其语义完整性与片段化策略直接影响文本向量化及后续检索效果。本研究在向量化原理的基础上，对文本进行分层清洗与结构化重构，通过保留关键语义单元、缩短chunks长度和明确上下文边界等方式，减少了在向量化过程中可能出现的语义稀释问题。该处理方式使文本表达更符合Transformer模型的注意力计算特性，有助于提高chunks召回率和准确率，为后续问答模块提供了可靠的数据基础。总结本次实践经验，笔者推荐各医疗机构在构建药学智能问答平台时部署Rerank模型对文本和向量检索结果进行tokens级别的语义相似性匹配与重新排序，以进一步提升检索准确性；如果未部署Rerank模型，选择语义检索与文本检索7∶3的混合策略，也能获得较为理想的召回效果。

4.2　平台构建策略与关键技术实现

本研究在平台构建过程中，运用Python自动化脚本完成了对院内1 120份药品说明书的批量清洗、格式统一与结构化重构，再将处理后的内容导入Dify平台。经结构化处理的文本在检索环节具有更高的匹配精度，结合Score阈值控制和问题分类器，可有效减少无关chunks的干扰，提升问答效率与准确性。

为确保知识来源的可控性与安全性，我院构建的药学智能问答平台依据本地部署模式运行，不具备自主学习功能，且与医院内部网络体系兼容。知识库缺项时会被系统自动记录并提示专业人员进行更新，从而保证内容的可追溯性和规范一致性。新增药品信息按最高优先级处理，相关资料在药品进院24 h内完成录入，满足了临床对药学信息的时效性需求。

本研究在研发时，同时对多种主流开源平台进行了比较：AnythingLLM在私有化部署方面具备优势[13]，但结果追溯与可视化能力有限；RAGFlow在复杂文档解析与引用可解释性方面表现突出[14]，但其硬件资源需求较高，限制了其在基层医疗机构的推广应用。综合部署成本、可视化程度和系统调试便利性，Dify平台在本研究所需的应用场景中具有更高的适配性。

4.3　平台运行效果与临床应用表现

为验证我院药学智能问答平台药学参数检索的准确性与可靠性，本研究随机抽取了我院20种非复方口服制剂，分别查阅其药品说明书中的t_{max}及$t_{1/2}$，并将药学智

能问答平台与在线版DeepSeek的检索结果进行对比。结果显示，我院药学智能问答平台给出的答案与药品说明书完全一致，准确率均为100%，而在线版DeepSeek的结果中，t_{max}准确率为30%，$t_{1/2}$准确率为50%，二者差异具有统计学意义（$P<0.001$），提示该平台在药学参数的检索和回答方面具有较高的准确性与可靠性。

在剂量个体化推理验证中，本研究针对12种需按肾功能调整剂量的药品对我院药学智能问答平台进行了分级测试，结果显示，所得剂量建议与药品说明书完全保持一致。而同一测试条件下，在线版DeepSeek的准确率为38%，两者差异具有统计学意义（$P<0.001$）。该结果提示，本研究构建的平台在处理涉及复杂条件判断的药学咨询时，仍然能够保持较高的可靠性。

在药物相互作用、禁忌证及特殊人群用药等高频临床查询场景中，该平台同样表现出良好的稳定性与准确度，不仅可通过工作流生成标准化回答，还能同步展示药品说明书原文，便于医师进行查阅与验证。此外，该平台还可与医院信息系统对接，实现全院范围的实时访问，提高了临床药学信息服务的可及性。

4.4 平台推广应用价值与可扩展性

基于标准化结构化知识库的本智能问答平台具有良好的可推广性，尤其适用于紧密型医疗集团的药学信息统一管理。通过共享同源知识库，可为紧密型医疗集团审方中心、辅助诊疗系统和语音问答工具提供一致且规范的知识支撑，有利于提升药学服务的同质化水平，并降低信息不一致带来的临床风险。

本研究将模型运行环境与知识库采用Docker镜像封装，实现了跨平台的一键迁移，减少了环境部署工作量，提高了平台可扩展性。相比依赖云端的方案，本地化部署具有更高的独立性和数据安全性，适用于网络条件有限或隐私保护要求较高的县域及基层医疗机构。

5 结语

LLM是否出现AI幻觉，主要取决于本地知识库是否存在相应的chunks以及其是否能被检索到。本研究构建的药学智能问答平台通过标准化重构药品说明书本地知识库，实现了高精度的检索与推理，大幅度减少了AI幻觉的出现。此外，通过集成提示词可以进一步优化LLM文字处理逻辑，如提示词添加“本地知识库没有找到相关资料，请联系静配中心药师进行维护”，既可以提醒药师不断完善知识库，亦可有效避免AI幻觉带来的临床决策误导。平台在上线之前与在线版DeepSeek进行了对比，结果显示，该平台几乎可以避免在DeepSeek上出现的AI幻觉问题，能够满足临床对药学信息严谨性与准确性的要求。本研究结果表明，通过知识库的标准化重构与轻量化本地部署，即便在有限算力条件下亦可实现可控、安全且质量可靠的药学智能问答服务。此类模式具有良好的普适性，适用于信息化建设需求迫切但算力资源受限的医疗机构，如基层医院、县域医疗中心及部分专科医院，能为其提供兼具成本-效益与专业性和可靠性的药学知识服务解决方案。

然而，本研究仍存在一定局限性，知识库的覆盖范围和更新频率仍需持续优化。未来研究可进一步扩大知识库规模，引入自动化更新机制，并探索多模态数据的整合，以进一步提升LLM在临床药学场景下的可靠性和实用性。例如，可进一步探索针对医学指南和专家共识的语义切分策略，并优化多模型协同检索与推理机制，以持续提升智能问答平台的性能。

参考文献

[1] 汪瑾，冷锴，陆慧. “互联网+”视域下智慧医疗服务模式创新研究[J]. 南京医科大学学报（社会科学版），2020，20（1）：84-87.

[2] 肖鹏，李新辰，朱斌，等. 基于互联网患者咨询的药师与人工智能回答一致性分析[J]. 临床药物治疗杂志，2025，23（5）：46-49，55.

[3] 国家卫生健康委办公厅. 国家卫生健康委办公厅关于印发卫生健康行业人工智能应用场景参考指引的通知[EB/OL]. （2024-11-14）[2025-08-09]. https://www.nhc.gov.cn/guihuaxxs/c100133/202411/3dee425b8dc34f739d63483c4e5c334c.shtml.

[4] PENG Y F，MALIN B A，ROUSSEAU J F，et al. From GPT to DeepSeek：significant gaps remain in realizing AI in healthcare[J]. J Biomed Inform，2025，163：104791.

[5] DREYER J. China made waves with DeepSeek，but its real ambition is AI-driven industrial innovation[J]. Nature，2025，638（8051）：609-611.

[6] 裴进涛，李昊东，刘广敏，等. 基于RAG的人工智能模型在高血压药品信息智能问答中的应用[J]. 电脑与信息技术，2025，33（5）：43-50.

[7] 查英华，郭朝霞，鞠慧光. 基于大语言模型的智能学习助手设计与实现[J]. 现代信息科技，2025，9（3）：50-55.

[8] 齐翌辰，王森淼，赵亚慧. 基于倒排索引的问答系统的设计与实现[J]. 教育教学论坛，2018（35）：68-69.

[9] DENG J W，ZUBAIR A，PARK Y J，et al. The use of large language models in medicine：proceeding with caution[J]. Curr Med Res Opin，2024，40（2）：151-153.

[10] ZHANG H J，DENG H R，OU J，et al. Mitigating spatial hallucination in large language models for path planning via prompt engineering[J]. Sci Rep，2025，15：8881.

[11] PAL A，UMAPATHI L K，SANKARASUBBU M. Med-MCQA：a large-scale multi-subject multi-choice dataset for medical domain question answering[EB/OL]. [2025-08-09]. https://arxiv.org/abs/2203.14371.

[12] CALIEBE A，LEVERKUS F，ANTES G，et al. Does big data require a methodological change in medical research?[J]. BMC Med Res Methodol，2019，19（1）：125.

[13] 李威威，夏新男. DeepSeek与Ollama协同的图书馆本地知识库构建[J]. 现代信息科技，2025，9（21）：148-154.

[14] 薛陶蓓，郑浩东. 基于DeepSeek和RAGFlow的智能项目推荐客服系统部署实践[J]. 现代信息科技，2025，9（18）：110-114.

（收稿日期：2025-08-13　修回日期：2025-12-24）

（编辑：孙　冰）