



# 献血咨询服务系统的设计与应用\*

——李大玮 陈委娜 单晓丽 刘鹏 焦淑贤

**【摘要】** **目的** 构建献血咨询服务系统,实现对标准化问题的自助解答。**方法** 基于检索增强生成技术,构建无偿献血专属知识库,依托大语言模型的自然语言对话能力,开发献血咨询服务系统。**结果** 系统运行4个月以来,血费报销和献血知识咨询量较同期减少5.37%和2.73%,表彰政策咨询量比上一个表彰年度减少28.89%。**结论** 基于大语言模型和检索增强生成技术构建的献血咨询服务系统,可以有效缓解人工客服工作压力,具有一定应用推广价值。

**【关键词】** 大语言模型;检索增强生成技术;无偿献血;献血咨询;献血服务

中图分类号:R197.6;R331.1

文献标识码:B

Research on the Design and Application of Blood Donation Consultation Service System/LI Dawei, CHEN Weina, SHAN Xiaoli, et al. // Chinese Health Quality Management, 2026, 33(3): 67-70

**Abstract** **Objective** To construct a blood donation consultation service system that enables self-service answers to standardized questions. **Methods** Based on retrieval-augmented generation technology, a dedicated knowledge base for voluntary blood donation was established. Leveraging the natural language dialogue capabilities of large language models, a blood donation consultation service system was developed. **Results** Since the system's operation for four months, the volume of inquiries regarding blood fee reimbursement and blood donation knowledge has decreased by 5.37% and 2.73%, respectively, compared to the same period. Additionally, inquiries about commendation policies have decreased by 28.89% compared to the previous commendation year. **Conclusion** The blood donation consultation service system, constructed based on large language models and retrieval-augmented generation technology, can effectively alleviate the workload of human customer service representatives and holds significant potential for broader application and promotion.

**Key words** Large Language Model; Retrieval-Augmented Generation Technology; Voluntary Blood Donation; Blood Donation Consultation; Blood Donation Service

**First-author's address** Qingdao Blood Center, Qingdao, Shandong, 266071, China

献血服务热线是连接采供血机构与献血者的纽带。以青岛市为例,2025年1-5月,青岛市中心血站献血服务热线共接收到献血者咨询电话9375人次,比2024年同期增长约60%。然而,受人工客服数量的限制,在应对献血咨询时存在排队现象严重、响应延迟等问题<sup>[1]</sup>。分析发现,咨询内容中,献血地点、检测结果、报销政策、献血知识等高频问题

占比达50%以上。此类问题若由人工客服处理,易造成人力压力与服务效率的双重挑战。为应对这一挑战,青岛市中心血站采用大语言模型和检索增强生成技术,构建了献血咨询服务系统,通过自动化处理献血知识、血费报销及表彰奖励3类标准化问题,不仅可以让人工客服聚焦于复杂咨询场景,还可以为献血者提供全天候服务,从而实现了服务资源的优

化配置。

## 1 系统架构

献血咨询服务系统包括大语言模型(large language model, LLM)、通用向量模型、向量数据库、无偿献血知识库、人机交互界面和图形工作站等,系统采用检索增强生成

DOI:10.13912/j.cnki.chqm.2026.33.3.15

\* 基金项目:青岛市2023年度医药卫生科研指导项目(编号:2023-WJZD188)

青岛市中心血站 山东 青岛 266071

(retrieval - augmented generation, RAG)技术实现。RAG 包含 3 个主要过程(图 1)<sup>[2]</sup>:检索是根据用户的查询内容,从外部知识库匹配相似性最高的数据进行排序;增强是将用户的查询内容和检索到的知识一起嵌入到一个预设的提示词模板中;生成是将经过检索增强的提示词内容输入到 LLM 中,以生成所需的输出。

### 1.1 知识库构建

整合《中华人民共和国献血法》《献血者健康检查要求(GB18467-2011)》等法律法规和标准要求,结合青岛市中心血站献血服务热线近 3 年的高频问答,汇总形成青岛市无偿献血知识库,涵盖献血注意事项、青岛市“四免”政策、血费减免政策及办理要求、优先用血及爱心献血注意事项等内容。采用关系型数据库构建知识库管理系统,系统提供标准化接口供通用向量模型调用,还可以依托青岛市智慧城市血液网实现与原献血服务热线的知识共享<sup>[3]</sup>。相比由文本构建的知识库,使用关系型数据库构建的知识库管理系统更新与维护更加方便、快捷,且支持对知识库的增量更新,避免了文本分段不合理导致的回答偏差<sup>[4]</sup>;同时,数据共享可以保证数据的一致性,减少了维护工作量,降低了知识库的管理门槛。

### 1.2 通用向量模型与向量数据库集成

在 RAG 技术框架下,通用向量模型负责将文本转化为向量表示。本研究中,青岛市中心血站使用了 BGE-Base 向量模型。BGE-Base 是一款通用型文本嵌入模型,支持处理长文本知识库。模型将数据库分块文档和用户提问转化为 768 维向量,提供给向量数据库进行存储

或检索。本研究中,青岛市中心血站使用 weavite 向量数据库。该数据库具备高效的向量检索能力、灵活的数据建模和多模态支持能力,能够快速构建基于语义的智能应用,并具备和 LLM 技术集成的能力。通用向量模型调用知识库数据,将其转化为向量,存储在向量数据库中。当用户提问时,通用向量模型接收用户提问信息,将其转化为向量,并向向量数据库中已存储的数据进行比对,实现向量检索<sup>[5]</sup>。

### 1.3 知识精炼与生成

LLM 是 RAG 技术系统的“智慧中枢”,负责接收检索增强后的信息并生成最终回答。LLM 可以根据实际需要和信息进行提纯并予以复杂的推理,然后将生成内容以对话形式输出。本研究中,青岛市中心血站选用的 LLM 是通义千问系列的 Qwen 2.5,相比 DeepSeek-R1, Qwen 2.5 具备更强的上下文能力以及更低的幻觉、更快的输出速度,可以给献血者提供更好的体验。为了更好地降低 LLM 的幻觉,提高输出的可靠性,需要通过提示词工程对 LLM 的输出进行优化<sup>[6]</sup>。提示词工程是通过设计、优化输入 LLM 的提示文本,引导模型生成更符合预期回答的关键技术(图 2)。

在 RAG 技术系统中,它架起了用户需求与模型输出之间的桥梁,直接影响回答的质量与实用性。

## 2 系统应用

献血咨询服务系统于 2025 年 3 月 28 日开始上线运行,于 6 月 25 日结合前期运行情况进行优化升级。据统计,2025 年 6 月 26 日-7 月 25 日期间,通过系统咨询记录达 946 条,咨询量排名前 5 的问题主要集中在奖励政策、血费报销和献血知识 3 个方面(表 1),占总咨询量的 40.59%(384/946)。从每天咨询时间段来看,非工作时间咨询量达 428 条,占总咨询量的 45.24%(428/946)。

对比系统上线 4 个月(2025 年 3 月 29 日-7 月 28 日)的人工客服数据,人工客服咨询记录为 8404 人次,比 2024 年同期增长 73.03%(3547/4857),其中非标准化问题占 74.89%,比 2024 年同期增长 30.52%;同期咨询血费报销和献血知识的数量分别减少 5.37%(19/354)和 2.73%(11/403)(表 2)。2025 年作为两年一度的全国无偿献血表彰年度,咨询表彰奖励政策的数量比上一个表彰年度(2023 年)减少 28.89%(562/1945)(表 3)。应用



图 1 献血咨询服务系统采用检索增强生成技术工作流程

```

# </context>
# When answer to user:
# - If you don't know, just say that you don't know.
# - If you don't know when you are not sure, ask for clarification.
# Avoid mentioning that you obtained the information from the context.
# And answer according to the language of the user's question.
# ""
self.template="""
请根据以下上下文内容,回答用户的问题。
<context>
{#{context#}}
</context>
注意:
1. 你作为青岛市中心血站的青岛市无偿献血智能客服助理,你的名字是“青小献”,你的主要任务帮助用户解答有关青岛市无偿献血的相关问题。
2. 首先判断用户问题与上下文内容的相关性。
3. 如果用户问题与上下文内容相关,请严格依据上下文内容和范围进行回答。
4. 如果用户问题与上下文内容无关或者无法从上下文找到相关内容,请明确告知用户你没有找到相关内容,你必须谨慎回答,不能编造和幻想。
5. 如果上下文中提到了96606工作电话,请在回答时严格依据上下文内容将工作时间与工作电话同时回答。
6. 请不要在回复中提及上下文和你是从上下文中获得信息的。
    
```

图 2 献血咨询服务系统提示词工程示例

表1 献血咨询服务系统咨询问题排名前5位统计

| 序号 | 咨询问题                   | 单位:条<br>咨询量 |
|----|------------------------|-------------|
| 1  | 献血多少可以办理献血荣誉卡          | 98          |
| 2  | 何为青岛市无偿献血“四免”政策        | 96          |
| 3  | 山东省“三免”政策与青岛市“四免”政策的区别 | 72          |
| 4  | 办理血费报销有哪些途径            | 60          |
| 5  | 献血有哪些注意事项              | 58          |

表2 献血咨询服务系统上线4个月血费报销与献血知识咨询人次比较 单位:人次

| 时间               | 咨询总量  | 血费报销   | 献血知识   |
|------------------|-------|--------|--------|
| 2024年3月29日-7月28日 | 4 857 | 354    | 403    |
| 2025年3月29日-7月28日 | 8 404 | 335    | 392    |
| $\chi^2$         |       | 68.144 | 72.083 |
| $P$              |       | <0.001 | <0.001 |

表3 献血咨询服务系统表彰奖励政策咨询人次与2023年度比较 单位:人次

| 时间       | 咨询总量  | 表彰奖励    |
|----------|-------|---------|
| 2023年    | 5 956 | 1 945   |
| 2025年    | 8 404 | 1 383   |
| $\chi^2$ |       | 513.773 |
| $P$      |       | <0.001  |

SPSS 27.0.1 统计软件进行数据分析,对3类标准化问题同期咨询量分别进行卡方检验, $P$ 均<0.05,差异具有统计学意义。

### 3 讨论

近年来,以LLM为代表的生成式人工智能的应用愈加广泛,尤其是DeepSeek发布以来,LLM应用呈现爆发式增长。随着应用的深入,LLM普遍存在的垂直领域知识准确性不足、信息不够全面等缺点逐步显现<sup>[7]</sup>。当前解决LLM垂直领域应用主要有3种方式,分别是对LLM进行训练、微调和构建专属领域知识库。虽然LLM训练和微调可以让其更好地理解专业领域的知识,但是通过构建专属领域知识库可以快速突破知识边界,降低LLM的训练成本,减少幻觉<sup>[8-9]</sup>,尤其在知识库内容发生变化时可以快速调整<sup>[10]</sup>。因此,构建专属领域知识库,通过LLM与RAG技术的结合突破LLM垂直领域知识局限性,是当前

生成式人工智能的重要技术方向。LLM与RAG技术的结合,在系统迭代效率和开发成本之间实现了平衡,同时保障了系统的准确率。在献血咨询服务系统立项之时,LLM的应用尚不广泛,初期也尝试采用了传统的规则匹配式自助问答服务,但在内部测试时发现,传统知识问答仅支持问题或关键词匹配,难以理解上下文,不能实现多步推理和复杂的语义理解,问题回答内容僵化,无法按照咨询人员的语境进行输出<sup>[11-12]</sup>。相比传统模式,在LLM与RAG技术的框架下,知识更新时效性、回答准确性、自然语言适应性以及可持续学习能力等方面有着天然优势。例如,当献血者提问“献血2 000 mL能否办理荣誉卡”时,传统模式无法给出答案,LLM则可以通过理解“办理献血荣誉卡必须献血4 000 mL以上的条件”得出不符合办理资格的结果。知识库则扮演着核心信息源和能力增强器的角色,其与LLM的结合突破了LLM的知识局限性,使模型既

能发挥语言生成优势<sup>[12]</sup>,又能依托外部知识库实现“知其然且知其所以然”,最终在准确性、时效性和专业性上实现质的提升。

本研究采用LLM和RAG技术构建献血专属知识库,为献血者提供无偿献血咨询服务。相比人工客服,献血咨询服务系统在回答标准化问题方面可以快速响应献血者需求,缩短排队时长,尤其是可以在非工作时间为献血者提供一站式问答服务。青岛市中心血站献血服务热线目前日常开放7个坐席,其中3个坐席保留用于日常献血招募工作,4个坐席用于献血者咨询,节假日开放1个坐席用于献血者咨询。一旦咨询献血者数量超过可用坐席数,献血者至少需要排队等待3~5 min,而使用献血咨询服务系统首次问题输出只需等待10 s左右,连续咨询输出等待时间只需3~5 s,且计算机系统并发访问数不再受制于人工坐席数,可以更好地支持多并发访问,大大缩短了标准化问题咨询等待时间。本研究通过系统上线后运行4个月的人工客服数据分析,在本年度人工客服咨询数量大幅度上升的情况下,血费报销和献血知识咨询量均比上一年度有所下降,咨询表彰奖励的人数比上一个表彰年度有了大幅度下降,说明系统应用对于人工客服有了较好的分流作用。而在使用该系统咨询的人员中,非工作时间咨询数量占比高达45.24%,说明系统应用为献血者在非工作时间提供了更加及时、便捷的服务。

但系统也存在一些问题:(1)献血者咨询过程随意性大,甚至出现错别字情况,导致LLM无法理解。针对这种情况,青岛市中心血站采用了两种方案予以优化:一是针对高频问题在咨询窗口置顶显示,如献血者需要可直接点击咨询;二是在献血者输入问题时,在问题输入

窗口按照输入内容给出联想问题,引导献血者规范提问。(2)知识库的覆盖度和关注角度问题。知识库作为系统的核心,高质量的知识库决定着问题回答的准确性。例如,青岛市中心血站在制订政策时,为了突出宣传献血者优待政策,知识库比较关注关于青岛市无偿献血“四免”政策的描述,但从献血者角度更加关注献血荣誉证书的办理。而在系统上线初期,知识库虽然包含了办理献血荣誉证书需要达到青岛市“四免”政策的要求,也包含了青岛市“四免”政策要求的条件,但是二者并不属于同一个知识块。由于负责匹配的通用向量模型并不具备推理能力,系统无法给出献血者所需的答案。针对这种情况,青岛市中心血站对知识库进行优化更新,设置了关联问题以解决这一问题。

综上,本研究通过LLM与RAG技术的融合,为献血咨询服务提供了智能化解决方案。实践表明,基于LLM和RAG技术的献血咨询服务系统可以有效缓解人工客服压力,为献血者提供更加及时、便捷的服务。虽然目前该系统还存在一些不足,但通过对知识库的不断优化更新以及未来LLM和通用向量模型能力的不断提升,系统将会更加

完善、精准,为献血者提供更加优质、高效的服务,从而推动无偿献血服务可持续发展。

**作者贡献:**李大玮负责方案设计、文章撰写;陈委娜负责数据分析、文献查阅;单晓丽负责知识库收集整理;刘鹏负责系统功能测试;焦淑贤负责方案指导与审核、终稿确认。

**利益冲突:**所有作者声明本文无实际或潜在的利益冲突。

#### 参考文献

[1] 裴君娜,孔长虹,金立明,等.献血咨询管理系统构建及应用[J].中国卫生质量管理,2022,29(7):94-96.

[2] 王亮.检索增强生成(RAG)驱动的知识服务:原理、范式及评估[J].科技与出版,2025(4):37-46.

[3] 李大玮,陈委娜,刘鹏,等.基于智慧城市概念模型智慧血液网建设与应用[J].中国卫生质量管理,2023,30(1):80-83.

[4] 高雅奇.基于大语言模型和RAG技术的高校知识库智能问答系统构建与评价[J].电脑知识与技术,2024,20(29):18-20,38.

[5] 文森,钱力,胡懋地,等.基于大语言模型的问答技术研究进展综述[J].数据分析与知识发现,2024,8(6):16-29.

[6] 吴璇,付涛.检索增强生成技术研究综述[J].计算机工程与应用,2025,61

(20):19-35.

[7] 姚畅,王贵宣,相鹏,等.医学大模型与多模态数据融合在智慧医疗中的机会、挑战与实践[J].中国卫生信息管理杂志,2025,22(2):171-178.

[8] 田雪晴,李泉江,游茂,等.我国医疗机构大语言模型建设现状调查与分析[J].中国卫生信息管理杂志,2025,22(1):38-44.

[9] BEDI P, THUKRAL A, DHIMAN S. XLR-KGDD: leveraging LLM and RAG for knowledge graph-based explainable disease diagnosis using multimodal clinical information [J]. Knowl Inf Syst, 2025 (prepublish): 1-21.

[10] 曹书林,史佳欣,侯磊,等.知识库问答研究进展与展望[J].计算机学报,2023,46(3):512-539.

[11] 贺李,刘兴红,贾鹏宇.基于RAG架构的智能知识库设计与应用研究[J].信息记录材料,2025,26(6):127-129.

[12] 富国瑞,王平利,王一展,等.基于大语言模型的高校图书馆智能参考咨询服务构建与应用研究——以山东大学图书馆为例[J].图书馆杂志,2025,44(12):31-40.

通信作者:

焦淑贤:青岛市中心血站副站长  
E-mail: jshuxian@sina.cn

收稿日期:2025-06-16

修回日期:2025-08-01

本文编辑:吴小红

## 本刊关于来稿中统计学处理的有关要求

作者应根据试验所采用的设计类型、资料所具备的特点和分析的目的,选择合适的统计分析方法。

文稿的统计学处理部分应写明所用统计方法的具体名称,如成组设计资料的 $t$ 检验、两因素析因设计的方差分析等。当涉及整体参数时,在给出显著性检验结果的同时,应给出95%可信区间。对于服从偏态分布的定量资料,应采用 $M(Q)$ 方式表达,不应采用 $\bar{x} \pm s$ 方式表达。对于定量资料,注意不盲目套用 $t$ 检验和方差分析。对于定性资料,注意不盲目套用 $\chi^2$ 检验。使用相对数时,分母不宜小于20。要注意区分百分率和百分比。统计学符号请按《统计学词汇与符号》(GB/T 3358.1-2009, GB/T 3358.3-2009)的有关规定书写,一律用斜体字母。

本刊编辑部