DOI：10.13998/j.cnki.issn1002-1248.26-0012

# 基于DeepSeek的健康中国战略下公共图书馆健康信息服务质量评估研究

韩　姝

（中共晋中市委党校，晋中 030606）

**摘　要：**[目的/意义]在“健康中国”战略背景下，针对公共图书馆健康信息服务评估维度单一、过程性证据缺失等问题，构建基于DeepSeek的多维评价体系并探索数智化评估路径。[方法/过程]以山西省32家公共图书馆为样本，融合问卷1 842份、服务日志约240万条与平台行为数据约560万条，结合专家访谈构建四维16项指标，并引入DeepSeek提取文本证据与风险线索。[结果/结论]评估结果显示，样本馆健康信息服务质量综合得分均值为3.43/5，整体处于中等水平。结构方程模型表明，专业度与易用性对整体质量的影响更显著，时效性与个性化次之，提示健康信息服务首先要守住可信与可用的底座。维度得分上，专业度与易用性相对较高，个性化得分最低且离散度最大，反映出馆际差异主要集中在分层入口、推荐机制与反馈闭环等环节。过程数据挖掘进一步发现，更新频率与满意度显著正相关（$r$=0.68），检索路径长度、二次检索率等指标可稳定识别服务卡点。干预验证采用前后对照准实验设计，在8个馆开展6个月改进，实验组综合质量提升17.5%，显著高于对照组3.2%，其中个性化与时效性改善更为明显。本研究的贡献在于把问卷感知、行为过程与文本证据对齐，形成可解释的“指标-证据”映射与整改清单，为公共图书馆健康信息服务的常态化评估、风险治理与精准改进提供方法与区域实证。

**关键词：**公共图书馆；健康中国；健康信息服务；DeepSeek；服务质量评估；数据挖掘

**中图分类号：**G250.2　　**文献标识码：**A　　**文章编号：**1002-1248（2026）10-0111-13

**引用本文：**韩姝．基于DeepSeek的健康中国战略下公共图书馆健康信息服务质量评估研究[J]．农业图书情报学报，2026，38(10)：111-123.

## 0　引　言

随着健康中国战略的深入实施，提升国民健康素养已成为国家的重要任务。公共图书馆作为社会知识与信息传播的核心平台，在健康信息服务中扮演着越来越关键的角色[1]。然而，当前国内公共图书馆的健康信息服务体系尚未完全适应新时代的需求，其服务质量评估方法也存在一定局限。例如，许多现有的评估体系依赖定量指标较少，评价维度过于单一，且数据分析能力尚显薄弱[2]。此外，传统的评估模型通常过度依赖问卷调查和主观评价，难以全面反映用户行为数据和隐性需求，导致评估结果缺乏准确性与时效性[3]。

随着人工智能技术的迅速发展，深度学习与数据挖掘技术的应用逐渐成为解决传统评估方法局限性的新途径。尤其是基于深度学习的模型，能够更好地挖掘数据中的潜在信息，并实现高效的数据分析。这一技术变革为提升公共图书馆健康信息服务质量评估的精准度和实时性提供了新的机遇。鉴于此，本研究拟

收稿日期：2026-01-08

作者简介：韩姝（1985-），女，本科，中共晋中市委党校，研究方向为图书馆学

DOI：10.13998/j.cnki.issn1002-1248.26-0012

构建基于DeepSeek的公共图书馆健康信息服务质量评估与数据挖掘模型，旨在通过融合多源数据、多维度评价与深度学习技术，对公共图书馆的健康信息服务质量进行更加科学和精准的评估与优化。具体而言，研究的创新点体现在3个方面：一是首次将DeepSeek深度学习技术应用于公共图书馆健康信息服务评估领域；二是构建了融合多源数据的评估指标体系；三是提出了基于数据驱动的服务质量优化路径。本研究不仅为提升公共图书馆在健康中国战略下的服务质量提供了理论支持，也为实际操作提供了实践指导，具有重要的理论和应用价值。

## 1　文献综述

健康中国战略持续推进，使公共图书馆健康服务由边缘探索逐步走向制度化议题。早期研究多从公共图书馆回应健康需求的必要性与合理性展开论证，强调健康服务在公共图书馆价值重估中的意义，并提出应将健康服务纳入图书馆服务规划与能力建设体系之中[1]。在此基础上，进一步的研究从价值认知与实践路径出发，结合国际组织、行业协会与公共卫生机构等多方视角，梳理公共图书馆健康服务的价值来源与可借鉴做法，提示健康服务的可行性不仅取决于资源供给，还依赖合作网络、组织机制与持续运营能力[2]。国内关于健康信息服务的经验总结亦逐步增多，研究从国内外实践归纳协作、培训与精准服务等要素，提出面向公共图书馆的健康服务发展框架，为后续评价指标的形成提供了服务链条与能力结构的参照[3]。

在用户研究层面，公共图书馆健康信息服务的满意度与行为意向成为常见切入点。基于顾客满意度相关模型的实证研究表明，期望、感知质量与感知价值等因素会影响满意度与持续使用意愿，提示健康信息服务应在可达性、体验一致性与价值兑现之间形成闭环[4]。考虑到老龄化背景下的服务公平问题，相关研究从老年人获得感出发，对公共图书馆适老化健康信息服务的类型与供给特征进行归纳，强调入口可见性、适老化呈现与线下支持的必要性，也提示不同人群对可理解性与操作负担的敏感差异[5]。与此同时，国外研究关于健康信息素养的述评显示，健康信息素养研究主题集中在图书馆角色、需求评估与多样化教育实践等方面，方法上偏重实证与跨机构合作，但公共图书馆场景的系统化评价仍存在不足[6]。

服务质量测量理论为图书馆健康信息服务评价提供了结构化工具。SERVQUAL量表提出的多维度感知质量框架为公共服务差距测量提供了经典路径，图书馆服务质量研究也长期借鉴其维度思想来构建指标体系[7]。在信息系统评价领域，信息系统成功模型及其更新版本强调信息质量、系统质量与服务质量对使用、满意度与净收益的影响，为将平台可用性、信息质量与服务过程指标纳入同一评价框架提供了理论支撑[8]。此外，图书馆服务质量测评工具“LibQUAL+”在高等教育与图书馆界被广泛应用，其研究综述指出“LibQUAL+”能够支持跨机构比较与持续改进，但也存在维度适配与解释边界等问题，需要结合具体业务情境进行本土化与场景化修订[9]。围绕“LibQUAL+”的心理测量学研究与维度结构讨论进一步表明，服务质量量表在不同用户群体中的维度稳定性与测量结构需要通过实证检验加以确认，这为在健康信息服务场景中强调可比性、稳定性与解释性提出了方法学要求[10,11]。

需要特别指出的是，健康信息服务不同于一般信息服务，其高风险属性决定了评价不能仅停留在体验与满意度层面。健康信息质量研究提出了多种可操作的核验工具与评价思路，其中DISCERN量表用于判断面向公众的治疗信息质量，强调依据、平衡性与可追溯性，反映了健康信息质量评价对可核验规则的重视[12]。针对网络健康信息质量的系统综述显示，不同研究在质量指标、测量口径与研究设计方面差异显著，且总体质量波动大、良莠不齐，提示需要建立更可复核的质量评价标准与治理机制[13]。从消费者视角出发的系统综述进一步指出，公众对网络健康信息的判断往往高度情境化且具有主观性，常用线索包括权威性、透明度、更新情况与界面体验等，但这些线索并不必然等同于医学准确性[14]。因此，在公共图书馆健康信息服务评价中，必须区分用户感知与医学核验，避免以可信感替代准确性判断，并将来源标注、更新机制与纠错流程纳入评价闭环。

近期研究开始讨论DeepSeek等工具在图书馆场景中的应用路径。在数字化转型背景下，多源数据驱动

评价成为提升评价客观性的方向。传统基于问卷与访谈的方法能够刻画感知与态度，但难以覆盖平台日志、检索路径、咨询响应与行为转化等过程性证据，进而影响对服务卡点的定位与改进效果的检验。将过程数据引入评价框架，有助于从可达性、摩擦成本与响应效率等维度解释满意度差异，也更利于形成可执行的整改清单。与此同时，面向非结构化文本的证据抽取逐渐成为提升解释性的关键环节。近期研究开始讨论DeepSeek等工具在图书馆场景中的应用路径，指出其可在战略规划、系统建设、运营管理与评价等层面提供支撑，但也强调应将技术能力转化为可解释、可追责的治理能力，避免停留在工具叙事。

综上，现有研究为公共图书馆健康信息服务的价值论证、服务框架、用户机制与评价工具提供了基础，但仍存在3方面需要进一步推进：第一，健康信息高风险属性下的准确性核验、来源追溯与纠错机制在图书馆评价体系中仍相对薄弱；第二，多源数据融合虽被倡导，但指标与证据的映射关系与解释性输出不足，难以把评价结果转化为具体改进动作；第三，用户异质性与场景差异对评价结果的影响需要更精细的分群刻画与验证。基于此，后续研究有必要在服务质量与信息系统评价框架的结构化基础上，引入可核验的健康信息质量约束，并通过多源过程证据与文本证据对齐，形成可解释的评估—改进闭环。

# 2 研究方法与评估指标

## 2.1 数据收集与处理

### 2.1.1 样本选择与数据来源

本研究采用分层抽样方法，选取山西省32家公共图书馆作为研究对象，确保样本的地域代表性与规模多样性。

（1）用户问卷。线上线下同步发放，回收有效问卷1 842份。问卷包括3类内容：用户基本信息；对健康信息服务的感知评价（采用5级量表，条目由文献整理并经专家审阅）；总体满意度与再次使用意愿。为减少一次性情绪影响，问卷引导被试以“最近一次健康信息需求经历”为参照作答。

（2）服务日志数据。来自32馆后台记录，涵盖检索词、借阅与下载、咨询工单、服务反馈等，合计约240万条。

（3）平台行为数据。来自数字平台的点击与停留行为，包含页面路径、停留时长、交互次数等，合计约560万条。

### 2.1.2 数据预处理

针对多源异构数据，本研究进行了系统化的预处理，确保数据质量与可分析性。

（1）剔除明显异常记录（如停留时长为0或极端值、重复写入、缺失关键字段等），并统一字段命名与编码规则。异常与缺失的处理原则以“不改变分布为先”，必要时采用分层抽样复核。

（2）对咨询问题与用户反馈进行分词、去噪与规范化，保留医学术语与常用缩写；对含义不明的短文本由两名研究者人工判读并形成标注规则，用于后续抽检。

（3）从过程数据提取可解释特征，例如：一次检索的迭代次数、从检索到获取目标资源的路径长度、咨询工单的首次响应时间、同类主题资源的覆盖度与更新间隔等。特征以周为单位聚合，既保留时序变化，又便于与问卷时间窗口对齐。初始特征共42项，后续通过相关性检验与共线性诊断进行裁剪。

（4）内容准确性抽检与高风险复核。考虑到健康信息错误代价高的问题，本研究不以用户主观感知替代医学准确性判断。对进入评估口径的健康信息内容（含专题条目、问答与咨询答复摘要）按主题分层抽样形成抽检集，由山西医科大学第一医院全科方向的副主任医师2名、山西医科大学第一医院药学方向的副主任医师2名及山西省J市卫生健康委公共卫生专家2名组成复核组完成核验；复核以权威指南、权威机构发布与可追溯来源为准，并形成差错记录与整改清单。依据权威指南、权威机构发布与可追溯来源进行核验，并记录差错类型（概念性错误、过期信息、适用人群不当、夸大或暗示性结论等）。对涉及用药、急症识别与就医决策的高风险文本，增加二次复核与来源追溯要求；凡无法提供可靠出处或表述存在歧义者，不纳入“内容准确性”加分，并在治理建议中形成整改清单。用户问卷中关于“准确”的评价仅用于反映用户对表述清晰度与可信感的感知，不作为医学准确性的直接证据。

DOI：10.13998/j.cnki.issn1002-1248.26-0012

## 2.2 DeepSeek模型应用与评估指标

### 2.2.1 DeepSeek技术框架

为避免将健康信息服务简化为问卷得分或单一算法输出，本研究采用“文本证据-行为证据-问卷感知”3类证据并行的评估思路。DeepSeek在框架中的定位并非替代所有算法，而是承担传统工具难以同时完成的3项关键工作：其一，将大量非结构化服务文本转化为可复核、可追溯的结构化证据；其二，将证据与指标建立稳定的映射关系，实现同口径融合；其三，对健康信息高风险内容进行识别与提示，使评估结果可以直接落实到治理与改进动作。

对比来看，规则或关键词法能够快速实现粗分类，但对同义表达、隐含意图、复合诉求与语境否定等情况鲁棒性不足；传统NLP（如TF-IDF/主题模型/浅层分类器）可用于统计性分析，但难以生成面向整改的原因要点与证据链；仅依赖BERT等判别式模型可以获得较好的分类指标，却仍难以在多任务下输出可解释的要点、完成指标映射与形成可执行的问题清单。DeepSeek的优势在于其指令式理解与结构化生成能力：在统一的任务协议与输出模板约束下，能够把主题、意图、原因要点、风险标签与证据片段稳定输出，并与指标体系对齐，形成面向改进的闭环数据。技术框架由3部分组成。

（1）文本证据抽取与语义结构化（DeepSeek核心环节）。数据来源包括线上咨询问题、留言反馈、活动评价与馆员答复摘要等。文本进入DeepSeek流水线后，先进行去标识化、分句与噪声清理，再在统一模板约束下完成3类任务，并输出结构化字段，供后续融合使用。①“主题-意图”识别：将文本映射到健康主题与服务意图标签，用于刻画高频需求与薄弱环节。该环节避免只按关键词命中，而是同时识别表达方式与请求目的，例如将同一主题下的不同诉求区分为资源检索、可信度核验、理解困难、更新滞后等。②原因要点抽取与倾向判别：对反馈文本提取负向原因要点及其证据片段，输出可直接进入整改清单的条目（如入口隐蔽、分类混乱、来源不明、更新过期、术语难懂、响应慢、流程烦琐等），并标注其对应的二级指标。与只给出正负向不同，本研究强调原因要点的结构化，是为了让低分项可解释、可行动。③指标映射与风险标注：DeepSeek按预定义规则将证据对齐到16项二级指标，形成“文本证据-指标映射”数据；同时对涉及用药、急症识别、诊疗决策相关内容加注高风险标签，进入更严格的规则过滤与人工复核流程。该环节的输出不是分值，而是带出处与片段的证据记录，便于复核与追责。需要说明的是，本研究所称的NLP分类器（BERT变体）用于支撑部分判别任务（如主题与情绪的稳定分类），属于DeepSeek文本流水线的底层组件之一；DeepSeek负责多任务编排、模板约束、结构化生成与指标映射，从而将分类结果提升为可解释证据与可执行清单。

（2）用户行为特征构建（过程证据，与DeepSeek互补）。从平台日志中提取与服务过程质量直接相关的可观测特征，包括检索次数与二次检索率、零结果检索占比、平均路径长度（从进入健康入口到到达目标内容的点击步数）、返回/退出比例、有效停留时长、主题页到详情页转化率、咨询提交到首次响应时间、闭环时间与重复咨询率等。行为特征用于刻画可达性、摩擦成本与响应效率等过程质量，避免用单一满意度替代服务过程的客观表现。行为证据与DeepSeek输出的文本原因要点互为校验：例如高零结果检索与找不到入口等原因要点可形成一致证据链，提高解释可信度。

（3）多源证据对齐、融合与评分输出（将DeepSeek输出变成可改进的评估结果）。将问卷条目得分、行为特征与DeepSeek生成的文本证据按同一指标口径汇总，得到4个一级指标与16个二级指标得分。融合阶段以加权汇总为主，同时保留解释层输出：对每个低分指标同步给出其行为证据（如零结果检索高、路径过长、响应超时）与文本证据（如分类混乱、入口隐蔽、来源不明、更新过期等），并生成按馆、按主题、按人群的整改线索清单。相较于传统的问卷均值汇报，这一机制使政策建议与优化措施能够被证据支撑，并可在干预前后复测中检验有效性。

### 2.2.2 评估指标体系构建与权重确定

（1）指标来源与形成路径。为避免指标设置停留在经验罗列，本研究采用文献与标准框架归纳、服务流程拆解、专家咨询收敛三阶段方法构建指标体系，并在每一步加入可测性与可治理约束。第一阶段，文

献与标准框架归纳。以服务质量与信息系统评价理论为基础，参考服务质量维度思想、信息系统成功模型对信息质量/系统质量/服务质量的划分，以及可用性与可访问性标准（ISO 9241-11:2018），并结合健康信息质量研究对准确性、权威性、更新、可理解性等要素的共识性要求，形成初始指标池。第二阶段，公共图书馆服务流程拆解。结合健康信息服务业务链条，将服务过程拆分为资源建设与组织、检索与导航、咨询响应与处理、热点专题更新、推荐与反馈闭环等环节，确保每项指标都对应可改进的业务动作与可采集的数据证据，并能在DeepSeek文本证据与日志行为证据中找到至少一种稳定观测口径。第三阶段，专家咨询收敛与可测性筛选。邀请公共图书馆信息服务、健康传播/公共卫生、信息系统与数据分析等领域专家开展两轮咨询，对指标的必要性、可操作性、可测性进行评分与合并，删除概念重叠或难以稳定测量的条目，最终形成指标。

（2）权重确定方法。权重采用“专家判断+层次分析法（AHP）”确定，步骤如下。①构建层次结构模型。目标层为“公共图书馆健康信息服务质量”，准则层为4个一级指标（PE、RC、SC、MM），方案层为16个二级指标。②形成判断矩阵。专家使用Saaty 1~9标度对同层指标进行两两比较，分别得到一级指标判断矩阵与各一级指标下二级指标判断矩阵。为降低个体偏差，采用多位专家判断结果的几何平均法合成群体判断矩阵。③计算权重并进行一致性检验。对判断矩阵求最大特征根及对应特征向量并归一化得到权重；一致性按AHP常规流程检验：

$$\text{一致性指标 } CI=(\lambda max-n)/(n-1) \quad (1)$$

$$\text{一致性比率 } CR=CI/RI \quad (2)$$

当$CR<0.10$时认为一致性可接受；若不满足，则返回专家复核并调整判断，直至通过。

（3）指标生成。基于以上步骤，得出公共图书馆健康信息服务质量评估指标体系与权重，详见表1。

### 2.2.3　数据挖掘与模型构建

本研究采用多种方法探索影响公共图书馆健康信息服务质量的关键因素，并在关键环节使用DeepSeek提升证据结构化与解释能力。

（1）用户分群。采用K-means与DBSCAN对用户群体进行分类，挖掘不同群体的服务需求特征。分群所用特征除行为变量外，还包含由DeepSeek抽取的主题—意图分布与原因要点特征，使分群不仅基于点击停留等外显行为，也能体现需求语义差异，从而提高分群的可解释性与可用于策略制定的价值。

（2）关联规则分析。使用Apriori算法挖掘需求类型与满意度之间的关联规则。规则项除问卷变量外，纳入由DeepSeek生成的原因要点标签与高频痛点，使关联规则能够直接指向可整改的原因类别，而不仅是抽象变量之间的相关性。

（3）结构方程模型。构建影响健康信息服务质量的路径模型，验证各评估维度间的作用关系。模型中部分观测变量来自DeepSeek证据统计（如某类原因要点出现率、某类主题投诉占比等），以增强潜变量解释的证据基础。

（4）融合评估模型。构建面向综合评分与质量等

表1　公共图书馆健康信息服务质量评估指标体系

Table 1　Evaluation index system for public library health information service quality

| 一级指标 | 二级指标 | 指标内涵（评估要点） |
|---|---|---|
| 专业度（PE） | 内容准确性 | 信息科学性、错误率、表述严谨性 |
| | 资源权威性 | 来源透明、权威出处、引用可追溯 |
| | 内容全面性 | 主题覆盖范围与深度、关键主题缺口 |
| | 更新及时性 | 内容发布时间与维护周期、过期内容占比 |
| 易用性（RC） | 检索便捷性 | 检索效率、检索结果可用性、零结果控制 |
| | 导航清晰度 | 分类逻辑清晰、路径短、入口可见性 |
| | 界面友好性 | 页面可读性、交互负担、表单与流程复杂度 |
| | 可访问性 | 无障碍与适老化支持、跨终端可用性 |
| 时效性（SC） | 响应速度 | 咨询受理到首次响应时间 |
| | 更新频率 | 专题与资源维护频次、更新间隔稳定性 |
| | 热点跟进 | 公共健康热点响应能力与专题上线速度 |
| | 处理效率 | 咨询闭环效率、重复咨询控制 |
| 个性化（MM） | 推荐精准度 | 推荐相关性、命中用户需求的比例 |
| | 需求适配度 | 分人群/分场景的内容供给与呈现适配 |
| | 交互便利性 | 提交咨询/反馈的便捷程度与引导清晰度 |
| | 反馈处理度 | 反馈受理、纠错改进与结果告知 |

DOI：10.13998/j.cnki.issn1002-1248.26-0012

级判别的融合模型。该模型的输入包含3类信息：问卷条目得分、行为特征与DeepSeek文本证据特征。与仅用BERT或仅用行为模型不同，本研究通过DeepSeek输出的结构化证据把文本与指标体系对齐，并保留解释项输出，使模型不仅可用于预测，也可用于定位短板并生成整改清单。

# 3 评估分析与结果

## 3.1 公共图书馆健康信息服务现状分析

### 3.1.1 服务内容结构与覆盖范围

对32家公共图书馆健康信息服务条目进行归类后，服务内容主要集中于健康科普、疾病防治与健康生活方式3类，医疗资源导航类占比较低（表2）。

从主题覆盖看，慢性病管理、营养健康、常见疾病预防是供给最集中的3个方向；精神健康、特殊人群健康、前沿医学研究相对薄弱。结合服务条目的组织方式与入口设置可见：薄弱并不完全等于“没有资源”，更常见的是缺少稳定入口、权威来源标注不足、阅读/使用路径不清晰，从而降低了可见度与可获得性。该特征与后文“易用性（RC）”“专业度（PE）”对综合评价的重要性一致。

表2 图书馆健康信息服务内容结构

Table 2 Service content structure of library health information services

| 服务内容类别 | 占比/% | 典型呈现形态 |
|---|---|---|
| 健康科普类资源 | 34.8 | 科普文章/短视频、主题书单、健康问答 |
| 疾病防治信息 | 28.3 | 常见病预防与康复、就医提示、风险识别 |
| 健康生活方式 | 18.6 | 饮食与运动、睡眠管理、控烟限酒 |
| 医疗资源导航 | 10.2 | 医院/科室/医保信息、挂号流程、机构链接 |
| 其他健康相关信息 | 8.1 | 政策与活动、公益讲座、健康日专题 |

为便于呈现主题覆盖的相对强弱，依据条目密度与入口可达性（是否有独立栏目、是否在检索推荐中可被触达）进行分级归纳（表3）。

表3 健康信息服务主题覆盖的相对分布

Table 3 Relative distribution of topic coverage in health information services

| 覆盖水平 | 主题领域（举例） | 主要表现 |
|---|---|---|
| 覆盖较高 | 慢性病管理、营养健康、常见疾病预防 | 条目量大、栏目稳定、检索命中率高 |
| 覆盖中等 | 运动健康、母婴与儿童基础保健、传染病基础知识 | 有内容但入口分散，更新不均衡 |
| 覆盖不足 | 精神健康、特殊人群健康、前沿医学研究 | 条目偏少或缺少权威标注，触达率低 |

### 3.1.2 DeepSeek平台用户需求结构与行为特征

用户健康信息需求既来自显性检索与咨询，也体现在停留、回访与路径迭代等隐性行为。本研究将日志会话与咨询工单主题进行合并归类，得到总体需求结构（表4）。常见疾病信息、健康生活指导、药物使用指南3类需求合计占比最高。

在“需求分群”层面，DeepSeek框架下的聚类结果（K-means与DBSCAN交叉验证）将用户划分为5类稳定群体（表5）。该结果用于后续差异化评价与优化策略定位：例如，“专业知识需求型”对权威来源与更新时效更敏感，而“健康监测管理型”对个性化与交互便利性更敏感。

在时效性与响应速度方面，问卷显示用户对“获得可用信息”的速度期望较高：76.5%的用户期望在检索后15分钟内获得可用信息；突发健康事件场景下的期望响应时间更短，均值为5.3分钟。在个性化方面，82.7%的用户希望获得更贴合自身状况的推荐与路径提示，尤其集中在慢病管理、健康饮食与特定人

表4 用户健康信息需求类型分布（按会话/工单主题归类）

Table 4 Distribution of users' health information needs (session/ticket based)

| 需求类型 | 占比/% |
|---|---|
| 常见疾病信息 | 32.6 |
| 健康生活指导 | 27.4 |
| 药物使用指南 | 15.8 |
| 其他（就医流程、检验解读、政策咨询等） | 24.2 |

表5　DeepSeek聚类识别的用户群体类型

Table 5　User segments identified by clustering under the DeepSeek framework

| 用户群体类型 | 占比/% | 主要需求特征（概括） |
|---|---|---|
| 健康知识普及型 | 36.2 | 常识科普、风险提示、可理解的解读 |
| 疾病信息查询型 | 24.8 | 症状与疾病对应、防治与就医建议 |
| 健康生活方式型 | 17.5 | 饮食运动、作息管理、体重与慢病预防 |
| 专业知识需求型 | 12.3 | 指南共识、研究进展、专业数据库入口 |
| 健康监测管理型 | 9.2 | 慢病随访、指标管理、个体化干预建议 |

群健康等主题。

## 3.2　服务质量评估结果

### 3.2.1　指标体系与权重说明

服务质量评估采用第二章构建的4个一级指标与16个二级指标体系（PE/RC/SC/MM），权重由AHP结合DeepSeek挖掘结果确定（表6为一级权重汇总，二级权重沿用表1）。

表6　一级指标权重

Table 6　Weights of the four primary dimensions

| 一级指标 | 记号 | 权重 |
|---|---|---|
| 专业度 | PE | 0.35 |
| 易用性 | RC | 0.27 |
| 时效性 | SC | 0.22 |
| 个性化 | MM | 0.16 |

通过DeepSeek挖掘确定的指标权重分布及指标间的相关性结构如图1所示。

在评估实现上，DeepSeek多源融合模块将问卷得分、日志特征（如检索迭代次数、路径长度、咨询首次响应时间、更新间隔等）与平台行为特征（点击、停留、回访等）进行注意力融合，形成综合评价输出。模型效果与传统方法对比详见表7，用于说明融合评估的稳定性与可用性。

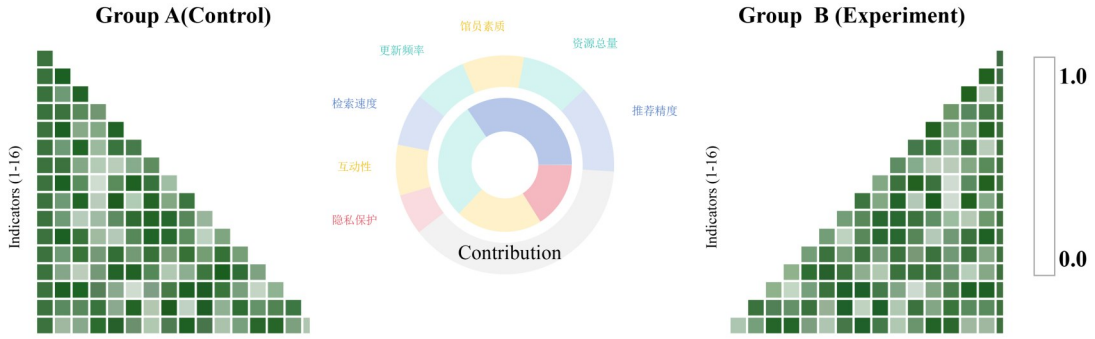



图1　健康信息服务评估指标体系权重构成与相关性结构

Fig.1　Weights and correlation structure of the evaluation indicator system

表7　DeepSeek关键模块的性能表现

Table 7　Performance of key modules in the DeepSeek-based framework

| 模块 | 任务 | 指标 | 结果 |
|---|---|---|---|
| NLP模块（BERT变体） | 反馈/咨询文本主题与情感识别 | F1 | 0.89 |
| 行为分析模块（MLP+LSTM） | 用户满意度预测 | 准确率 | 82.40% |
| 多源融合模块（注意力融合） | 综合评分回归 | RMSE | 0.086 |
| 线性融合（对照） | 综合评分回归 | RMSE | 0.142 |
| 深度学习综合判别 | 质量等级判别 | 准确率 | 87.60% |
| 传统机器学习（对照） | 质量等级判别 | 最高准确率 | 76.80% |

DOI：10.13998/j.cnki.issn1002-1248.26-0012

#### 3.2.2 总体评估结果与维度差异

对32家公共图书馆计算综合得分后，总体均值为3.43分（满分5分）。4个一级指标均值与离散程度详见表8。整体呈现“专业度、易用性相对较高，个性化相对较弱且差异更大”的特征。

表8 图书馆健康信息服务质量得分（1~5分）

Table 8 Scores of health information service quality（1~5）

| 指标维度 | 均值 | 标准差 |
|---|---|---|
| 专业度（PE） | 3.62 | 0.43 |
| 易用性（RC） | 3.52 | 0.38 |
| 时效性（SC） | 3.38 | 0.51 |
| 个性化（MM） | 3.19 | 0.57 |
| 综合得分 | 3.43 | - |

#### 3.2.3 不同类型图书馆的评估比较

按规模与功能特征将32馆分为4类进行对比：大型/省级馆、中型/地市级馆、县区级馆、特色馆。各类在4个维度上的表现详见表9（本研究按原始统计报告口径呈现已明确的数值项）。

从对比结果看，“资源优势”并不必然带来“过程优势”：大型馆的PE领先明显，但SC并未同步拉开差距。特色馆在MM上优势突出，提示机制设计（分层入口、规则推荐、互动闭环）对个性化有直接贡献。

#### 3.2.4 不同用户群体的评价差异

在用户评价层面，DeepSeek聚类后的群体差异更清晰：不同群体对4个维度的权重感知并不一致。综合问卷结果，差异主要体现在易用性与专业度，以及对时效与个性化的敏感程度不同。典型群体的维度均值详见表10。

表9 不同类型图书馆健康信息服务质量得分对比（均值，1~5分）

Table 9 Comparison of service quality scores by library type（mean，1~5）

| 类型 | 专业度（PE） | 易用性（RC） | 时效性（SC） | 个性化（MM） | 主要特征概括 |
|---|---|---|---|---|---|
| 大型/省级馆 | 4.08 | - | 3.42 | - | 专业资源集中，响应与更新链条较长 |
| 中型/地市级馆 | 3.28~3.56 | 3.28~3.56 | 3.28~3.56 | 3.28~3.56 | 各维度相对均衡但整体中等 |
| 县区级馆 | 3.15 | 3.63 | - | - | 流程较简洁，专业资源相对不足 |
| 特色馆 | - | - | - | 3.82 | 个性化机制突出，维度间差异较大 |

*注：表内数值为同口径统计结果。对样本量不足或不具可比性的维度，不报告，用“-”表示

表10 典型用户群体的维度评价差异（均值，1~5分）

Table 10 Differences in dimension ratings across typical user groups（mean，1~5）

| 群体 | 易用性（RC） | 专业度（PE） | 时效性（SC） | 个性化（MM） | 评价侧重点 |
|---|---|---|---|---|---|
| 儿童与老年用户 | 4.26 | - | 3.12 | - | 更看重“好找、好懂、少步骤” |
| 医疗专业背景用户 | - | 3.85 | 3.67 | - | 更强调“权威来源、更新速度” |
| 教师群体 | - | 3.73 | - | 3.43 | 兼顾教育价值与推荐能力 |

方差分析结果表明，不同群体在4个一级指标上的评分差异达到显著水平（$p<0.01$），说明以“单一入口+同一呈现方式”提供健康信息服务难以覆盖主要人群，需要在入口组织与内容呈现上进行分层。

### 3.3 基于DeepSeek的前后对照干预效果

为检验评估框架的可改进性，本研究采用前后对照的准实验设计。从32馆中选择8馆进入干预验证，选择原则为：平台形态与数据记录较完整、基线综合得分处于中等区间且具备改造空间，同时覆盖不同规模类型以提高结果解释力。分组遵循同类可比原则，先按馆舍规模、服务量与干预前综合得分进行配对匹配，再在配对内确定实验组4家、对照组4家，以降

低资源禀赋差异带来的偏差。干预周期为6个月：第1个月完成问题清单确认、责任分工与上线计划；第2~5个月实施改造并按周记录更新、响应与推荐等关键过程指标；第6个月完成后测评估与抽检复核。

实验组干预措施与指标一一对应。专业度方面，统一来源标注模板并建立内容抽检与高风险复核流程；易用性方面，重构健康入口与主题聚合页，针对高频零结果检索完善同义词表、纠错与替代资源提示；时效性方面，设定咨询首次响应与工单闭环时限，建立热点专题快速上架流程；个性化方面，先以规则实现分层入口与推荐，再引入基于用户分群的路径提示与反馈闭环。对照组维持原有流程，仅进行常规运维，不实施上述针对性改造。前测与后测均使用同一指标口径与数据处理流程，并通过差异检验评估干预效果。

本研究选取8家典型公共图书馆开展实证研究。4家为实验组、4家为对照组。实验组依据DeepSeek评估结果实施优化，主要包含4类措施：专业资源整合与权威标注、用户界面与导航优化、更新机制改进、个性化推荐与分层入口升级。干预周期为6个月。

干预前后对比结果详见表11。实验组服务质量平均提升17.5%，显著高于对照组的3.2%（$p$<0.001）；4个一级指标中，个性化（MM）提升最明显（+24.8%），其次为时效性（SC）（+18.3%）。满意度方面，实验组由3.46提升至4.12，而对照组仅由3.42提升至3.53。

表11 干预前后评估结果对比（实验组vs对照组）

Table 11 Pre-post comparison of intervention results (experimental vs. control)

| 指标 | 实验组（前→后） | 变化幅度/% | 对照组（前→后） | 变化幅度/% |
|---|---|---|---|---|
| 综合服务质量 | - | 17.50 | - | 3.20 |
| 个性化（MM） | - | 24.80 | - | - |
| 时效性（SC） | - | 18.30 | - | - |
| 总体满意度 | 3.46→4.12 | 66.00 | 3.42→3.53 | 11.00 |

在描述完干预实验的总体结果后，进一步对各细分指标的干预效应进行分析（图2），发现推荐精准度与响应速度的改善最为显著。

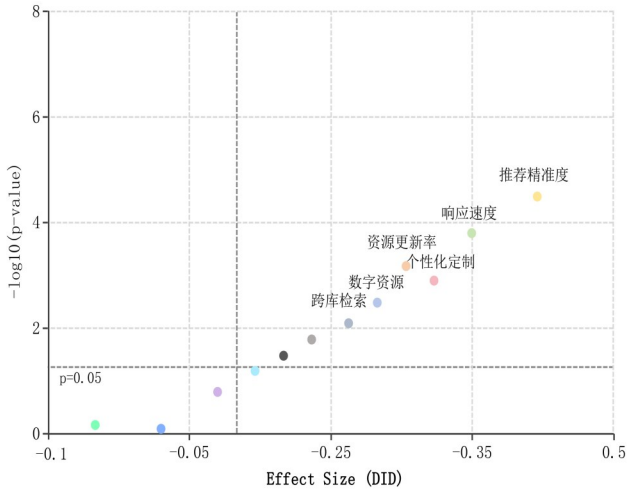



图2 干预实验前后各服务指标改善效果分析

Fig.2 Improvements in service indicators before and after the intervention

分群体效果呈现差异。儿童与老年用户对易用性改进的满意度提升最明显（31.2%）；医疗专业背景用户对专业度提升的认可度更高（26.8%）。该结果说明，基于DeepSeek的多源评估不仅能够识别短板，还能将优化落实到“对谁有效、因何有效”的层面，从而提高改进的针对性。

## 3.4 影响服务质量的关键因素分析

### 3.4.1 结构方程模型结果

为检验指标体系的结构合理性与作用路径，本研究构建结构方程模型。模型拟合良好：$\chi^2/df$=2.48，CFI=0.943，TLI=0.931，RMSEA=0.058。4个一级维度对整体服务质量的标准化路径系数见表12，均达到显著水平（$p$<0.001）。

结果表明，提升整体质量首先要守住“专业可信”和“易获取易理解”的底座，其次才是响应速度与个性化能力的进一步强化。

DOI：10.13998/j.cnki.issn1002-1248.26-0012

表12　结构方程模型标准化路径系数

Table 12　Standardized path coefficients of the structural equation model

| 路径 | 标准化系数$\beta$ | 显著性 |
|---|---|---|
| 专业度（PE）→服务质量 | 0.42 | $p$<0.001 |
| 易用性（RC）→服务质量 | 0.37 | $p$<0.001 |
| 时效性（SC）→服务质量 | 0.28 | $p$<0.001 |
| 个性化（MM）→服务质量 | 0.24 | $p$<0.001 |

#### 3.4.2　DeepSeek挖掘支持的关键因素

在过程数据层面，DeepSeek对日志与行为特征进行建模与对齐后，识别出与满意度和综合得分关联最稳定的因素，归纳为5类（表13）。其中，“更新频率”与满意度呈显著正相关（$r$=0.68，$p$<0.001）。“界面与检索友好性”与完成路径长度、二次检索率等行为指标密切相关。“专业性与可理解性平衡”在反馈文本中高频出现，是影响服务可用性的关键环节。

表13　影响服务质量的关键因素（融合证据汇总）

Table 13　Key factors influencing service quality（evidence summary）

| 关键因素 | 数据证据来源 | 主要结果（按本研究统计口径） |
|---|---|---|
| 信息更新频率 | 日志特征+满意度 | 与满意度正相关$r$=0.68，$p$<0.001；事件期作用更突出 |
| 检索系统用户友好性 | 行为路径特征 | 简洁入口与更短路径对应更高满意度（提升幅度在多馆对比中稳定出现） |
| 专业性与可理解性平衡 | NLP主题/情感+访谈 | “看不懂”“术语多”“缺少解释”与低评分高度同现 |
| 个性化推荐精度 | 推荐点击/回访+问卷 | 推荐更贴合时，满意度显著上升（与MM维度一致） |
| 社区互动性 | 互动次数+反馈闭环 | 互动与评价提升相关（$\beta$=0.32，$p$<0.01） |

## 4　研究讨论

### 4.1　DeepSeek与公共图书馆健康信息服务质量评估的理论意义

公共图书馆健康信息服务研究长期面临一个方法论张力：一方面，服务质量测量强调以结构化维度刻画用户感知，并将其作为持续改进的依据[7,9]；另一方面，健康信息具有高风险属性，单靠感知评价难以确保内容可靠，且用户用于判断的线索并不必然等同于医学准确性[12-14]。本研究引入DeepSeek构建多源证据融合框架，其理论意义首先体现在对这一张力的“可操作化调和”：在SERVQUAL、信息系统成功模型等既有框架提供的维度结构基础上[7,8]，把健康信息质量领域强调的可追溯性、平衡性与核验要求嵌入评估流程[12,13]，使服务质量评价不再仅是体验测量，而成为兼具治理约束与改进导向的证据体系。

其次，本研究推进了图书馆服务质量评价从“结果性测量”向“过程性解释”的理论转向。既有研究指出，“LibQUAL+”等工具能支持跨机构比较与持续改进，但在具体业务情境中仍存在解释边界与适配问题，需要进行场景化修订[9-11]。健康信息服务的关键差异在于：同样的满意度得分背后可能对应着完全不同的风险结构与过程摩擦。本研究通过DeepSeek将咨询与反馈文本抽取为“主题-意图”与原因要点，并映射到二级指标，使每个维度得分都对应可复核的证据集合，从而把维度结构由“统计量”推进为“可解释的证据构成”。这一做法强化了测量的内容效度与可治理性，也为解释为何专业度与易用性成为质量底座提供了理论支撑：在健康情境下，专业度对应信息质量底线，易用性对应系统质量与可达性，两者共同决定了使用与满意度的基础条件，而个性化更体现为运营机制成熟度与闭环能力差异[3,5]。

再次，从公共图书馆健康服务价值与实践研究看，健康服务的可持续性不仅取决于资源供给，更取决于组织机制与协作网络[2,3]。本研究的贡献在于把“组织机制”以可测证据形式嵌入评价体系：更新频率、响应链条与反馈闭环等过程变量通过日志与文本证据被同时观测，并可用于解释满意度与质量差异。这使健康信息服务的质量评价从经验性描述走向可复核的治理研究路径，也回应了既有研究关于影响因素识别与实践框架构建的需求[1]。综上，DeepSeek在本研究中的理论意义不在于替代传统量表，而在于使服务质量维度、健康信息质量约束与过程证据对齐成为

可能，从而为公共图书馆健康信息服务的学术评价提供一种更适配高风险内容的解释框架。

## 4.2 DeepSeek介入的理论机制解释

传统服务质量框架强调从用户感知出发构建维度，并据此衡量差距与改进方向[7,9]；信息系统成功模型则指出信息质量与系统质量通过使用与满意度影响最终收益[8]。但在健康信息情境中，服务对象的风险感知更强，用户对可靠性、权威性与可追溯性的关注更集中，且用户用于判断信息质量的线索未必等同于医学准确性[12-14]。因此，评价不仅要回答用户“满意不满意”，还要回答服务“为什么被认为可靠或不可靠”，并能在同一理论维度下形成可解释的比较。

DeepSeek介入后，最关键的机制变化在于“意义的归并”与“概念的落地”。一方面，健康信息服务中大量问题以自然语言叙述的方式出现，既包含事实，也包含情绪、期待与责任归属；若仅依赖问卷条目，容易把复杂经验压缩为单一分值，导致维度得分缺少解释力。DeepSeek使这些叙述得以在同一语义口径下被归并为可讨论的服务现象，例如把入口隐蔽、分类混乱、来源不明等表达归入易用性或专业度的不同子维度，从而将抽象构念转化为可被讨论的具体内涵。这种归并并不是技术上的分类，而是理论上的概念化过程，使服务质量维度成为对现实问题的有效解释框架[10,11]。

另一方面，健康信息质量研究强调依据、平衡性与可追溯性[12]，并指出网络健康信息质量波动大、评价标准不一[13]。DeepSeek的理论贡献在于把这些质量约束嵌入服务质量维度之中，使专业度不再只是“感觉专业”，而成为与来源透明、更新逻辑、风险提示等治理要素相联结的评价维度。这一转变使服务质量评价从体验导向走向治理导向：既保留用户感知作为公共服务改进的重要依据[4,5]，又避免在高风险情境下以可信感替代可靠性判断[14]。因此，DeepSeek在本研究中更像一种“解释框架的增强器”，它强化了质量维度与实际服务经验之间的对应关系，使评估结果能够承载理论解释而不仅是结果陈述，并为后续改进提供更清晰的因果线索与优先级。

## 4.3 实践启示

公共图书馆健康信息服务的改进，确实可以从常识层面提出“要专业、要好用、要及时、要个性化”。但本研究基于DeepSeek的意义在于：把这些看似正确的方向落到可执行、可追踪、可复盘的治理单元上。DeepSeek将咨询与反馈文本抽取为主题—意图与原因要点，并与日志与行为证据对齐，形成3类直接产出：一是按指标归类的高频痛点清单；二是按主题与人群分层的需求画像；三是高风险内容与过期信息的提示与整改清单。下述建议均由这些产出支持，并提供可监测的改进抓手，避免停留在口号上。

（1）建立可核验的专业底线。建议将专业度治理从“要求馆员把关”转为“问题清单驱动的内容治理”。依托DeepSeek输出的高风险主题识别与原因要点（如来源不明、表述含糊、适用人群不清等），建立两张清单：第一张是高风险内容清单，集中覆盖用药、急症识别、诊疗决策相关内容，实行更严格的复核与上线门槛；第二张是可疑内容清单，覆盖过期信息、出处不全、结论夸大等问题，要求补齐来源标注或下线替换。两张清单都应当能回溯到原文片段与出处，形成可问责的证据链。

（2）促进DeepSeek的易用性。易用性问题往往不是“界面不好看”，而是用户在关键步骤上反复受阻。DeepSeek抽取的原因要点中，入口隐蔽、分类混乱、找不到、检索无结果等属于高频、可行动问题；这些问题与日志中的二次检索率、零结果检索占比、路径长度、回退退出等特征可以形成一致指向。建议以此建立健康信息服务的卡点地图：把问题按入口、检索、导航、阅读、咨询5个环节拆开，逐项给出整改动作（入口固定化、同义词表与纠错提示、主题聚合页、替代资源与一键咨询等），并按周复盘。

（3）时效性不只看响应速度，而要用“事件-主题”驱动的更新机制替代临时宣传。更新频率与满意度的相关性表明，用户把更新视为可靠服务的信号。传统建议往往停留在“提高更新频率”，但缺少抓手。DeepSeek可从咨询与反馈中识别季节性与事件性主题的集中出现（如流感季、慢病随访、饮食管理等），并形成主题热度清单。

DOI：10.13998/j.cnki.issn1002-1248.26-0012

（4）优化个性化分层服务。本研究分群结果表明不同人群对专业度、易用性、时效性与个性化的敏感点不同。个性化短板的根源往往是缺少分层入口与反馈闭环，而非缺少算法。建议用DeepSeek形成的“人群-主题-意图”画像，把健康信息入口做成分层结构：至少按老年友好、慢病管理、母婴与儿童、心理健康等场景提供稳定入口，并在每个入口内设置少量高频任务的快捷路径（例如我想查什么、我该去哪里、我需要咨询等）。在推荐上先采用规则与场景匹配，待标签与反馈口径稳定后，再引入更复杂的模型推荐。

（5）推进数据治理保障常态化评估。DeepSeek的不可替代性还体现在“可持续复盘”：它能把日常产生的海量文本转成结构化问题与趋势线，使评估不是一年做一次的项目，而是每月都能运行的机制。建议建立健康信息服务的月度体检制度：固定输出4类报表——高风险内容清单、痛点原因要点TOP榜、主题热度与需求迁移、关键过程指标（检索/路径/响应/闭环）。并明确数据口径与权限管理，确保不同馆之间可比、过程可审计。

## 5 结 语

本研究围绕公共图书馆健康信息服务质量，构建了由专业度、易用性、时效性、个性化4个一级指标与16个二级指标组成的评估体系，并以问卷、服务日志与平台行为数据进行联合验证。32家图书馆的综合得分均值为3.43分（满分5分），整体处于中等水平。4个维度中，专业度均值3.62、易用性3.52，相对较高；时效性3.38居中；个性化3.19最低且离散度最大，说明馆际差异主要集中在推荐、分层入口与反馈闭环等环节。二级指标中，内容准确性得分最高（4.12），推荐精准度最低（2.87），反映出多数馆能够守住内容基本质量，但在信息触达与适配方面能力不足。结构方程模型结果显示，专业度与易用性对整体质量的影响更强（PE $\beta$=0.42，RC $\beta$=0.37），时效性与个性化次之（SC $\beta$=0.28，MM $\beta$=0.24），提示提升服务应优先巩固可信与可用的基础。过程数据挖掘进一步指出，更新频率与满意度显著正相关（$r$=0.68），检索路径长度、二次检索率等行为指标可稳定识别服务“卡点”。干预实验表明，依据评估结果实施针对性改进后，实验组综合质量提升17.5%，显著高于对照组3.2%，其中个性化与时效性提升最明显，且不同人群对改进项的响应存在差异。总体看，多源数据评估能够将主观感受与过程证据对齐，为定位短板与检验改进提供了可靠依据。

当然，本研究也有不足的地方。本研究以山西省32家公共图书馆为样本，结论主要反映中部省份公共图书馆在健康信息服务建设与数字化治理阶段的现实特征。不同地区在资源供给能力、平台形态、人口结构与协作网络上存在差异，指标权重与干预效果的外推需结合地方条件审慎解释。后续研究将通过跨区域样本与分层抽样检验指标体系与模型参数的稳定性，并比较不同地区在专业度、时效性与个性化机制上的差异，以提升研究结论的可迁移性。

## 参考文献：


[1] 王国凯. 基于模糊DANP的公共图书馆健康信息服务质量影响因素识别研究[J]. 农业图书情报学报, 2022, 34(10): 33-43.
Wang Guokai. Identification of influencing factors of health information service quality in public libraries based on fuzzy DANP[J]. Journal of Library and Information Science in Agriculture, 2022, 34(10): 33-43.

[2] 周晓英, 于小宁. 公共图书馆健康服务的价值认知和实践探索[J]. 中国图书馆学报, 2022, 48(4): 90-103.
Zhou Xiaoying, Yu Xiaoning. Value cognition and practice exploration of public library health service[J]. Journal of Library Science in China, 2022, 48(4): 90-103.

[3] 刘聪. 公共图书馆健康服务框架研究——基于国内外健康服务实践的分析[J]. 图书与情报, 2020(4): 92-100.
Liu Cong. Public library health service research - Analysis based on domestic and foreign health service practice[J]. Library and Information, 2020(4): 92-100.

[4] 倪伟宸. 公共图书馆健康信息服务用户满意度影响因素研究[J]. 图书馆研究与工作, 2023(2): 16-24.
Ni Weichen. A study on the factors influencing user satisfaction of health information services in public libraries[J]. Library Science Research & Work, 2023(2): 16-24.

[5] 刘一鸣, 吴壮波. 基于老年人获得感的公共图书馆适老化健康信息服务研究[J]. 图书馆研究与工作, 2023(5): 5-12, 57.
Liu Yiming, Wu Zhuangbo. Research on elderly-oriented health information services in public libraries based on a sense of gain of the elderly[J]. Library Science Research & Work, 2023(5): 5-12, 57.

[6] 唐鹏宇. 国外图书馆健康信息素养研究述评[J]. 图书馆研究与工作, 2021(2): 37-42.
Tang Pengyu. A research review of health information literacy in foreign libraries[J]. Library Science Research & Work, 2021(2): 37-42.

[7] Parasuraman A, Zeithaml V A, Berry L L. SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality[J]. Journal of Retailing, 1988, 64(1): 12-40.

[8] Delone W H, McLean E R. The DeLone and McLean model of information systems success: A ten-year update[J]. Journal of Management Information Systems, 2003, 19(4): 9-30.

[9] 豆淑云. LibQUAL+图书馆服务质量评价研究综述[J]. 知识管理论坛, 2013(9): 45-51.
Dou Shuyun. A summary of service quality evaluation of LibQUAL+ library[J]. Knowledge Management Forum, 2013(9): 45-51.
[10] Cook C, Thompson B. Psychometric properties of scores from the web-based LibQUAL+ study of perceptions of library service quality[J]. Library Trends, 2001, 49(4): 585-604.
[11] Thompson B, Cook C, Heath F M. How many dimensions does it take to measure users' perceptions of libraries?: A LibQUAL+Study[J]. portal: Libraries and the Academy, 2001, 1(2): 129-138.
[12] Charnock D, Shepperd S, Needham G, et al. DISCERN: An instrument for judging the quality of written consumer health information on treatment choices[J]. Journal of Epidemiology and Community Health, 1999, 53(2): 105-111.
[13] Eysenbach G, Powell J, Kuss O, et al. Empirical studies assessing the quality of health information for consumers on the world wide web: A systematic review[J]. JAMA, 2002, 287(20): 2691-2700.
[14] Sun Yalin, Zhang Yan, Gwizdka J, et al. Consumer evaluation of the quality of online health information: Systematic literature review of relevant criteria and indicators[J]. Journal of Medical Internet Research, 2019, 21(5): e12522.

# Health Information Service Quality Evaluation of Public Libraries under the Healthy China Strategy Using DeepSeek

HAN Shu
(Party School of the CPC Jinzhong Municipal Committee, Jinzhong 030606)

**Abstract:** [Purpose/Significance] As part of the Healthy China strategy, public libraries are expected to serve as accessible hubs for health knowledge dissemination and community health literacy improvement. However, health information services are inherently high stakes: users' perceived credibility does not necessarily guarantee medical accuracy, and outdated or poorly contextualized information may lead to serious consequences. Evaluations in existing library settings still rely heavily on surveys and experience-based summaries. While these are useful for describing satisfaction, they are often insufficient for identifying process bottlenecks, explaining why specific dimensions perform poorly, or supporting the accountable governance of high-risk content. This study aims to develop an evidence-oriented and interpretable evaluation approach for public library health information services, and to demonstrate how DeepSeek can enhance the measurability and actionability of quality assessment beyond conventional questionnaire-and-interview approaches.[Method/Process] We selected 32 public libraries in Shanxi Province for our empirical study. A four-dimension, sixteen-indicator framework was established (Professionalism, Ease of Use, Timeliness, and Personalization) through literature consolidation, service-process decomposition, and expert consultation. Multi-source data were collected and aligned, including 1 842 valid questionnaires, approximately 2.4 million service log records, and about 5.6 million platform behavior records. DeepSeek was employed to transform unstructured service texts (online consultations, user feedback, activity reviews, and staff response summaries) into structured evidence: it consolidates theme–intent patterns, extracts reason-oriented points that explain dissatisfaction, and maps those textual cues to specific secondary indicators. In parallel, process features such as zero-result search rate, path length, response time, and repeated consultation rate were derived from logs to represent observable friction and efficiency. An intervention validation was conducted using a quasi-experimental pre-post design with matched experimental and control libraries over six months, applying evidence-driven improvements linked to low-performing indicators. [Results/Conclusions] The overall mean score of health information service quality was 3.43 out of 5, indicating a medium level across the sample. Structural analysis shows that professionalism and ease of use exert stronger effects on overall quality, implying that trustworthy content governance and accessible service pathways form the foundational layer of health information services. Personalization receives the lowest average score and the largest between-library variation, suggesting that differentiated entry design and feedback closure mechanisms are the main sources of divergence. Update frequency is positively associated with satisfaction, highlighting the practical importance of routine maintenance rather than campaign-style updates. In the intervention study, the experimental group achieved a 17.5% improvement in overall quality, significantly higher than the control group. The findings indicate that aligning perceived evaluations with behavioral evidence and text-based explanatory cues can improve both interpretability and implementability of quality assessment. Limitations include regional sampling and potential differences in logging completeness. Future work should expand cross-regional validation and incorporate outcome-oriented indicators that more accurately reflect problem resolution.

**Keywords:** public library; healthy China; health information services; DeepSeek; service quality assessment; data mining