



# 高校图书馆智慧服务效能动态多维评价体系构建与实践

杨梦洁 侯点可 倚海伦 黄辉 张盛兵\*

**摘要** 高校图书馆服务效能评价旨在引导图书馆不断满足用户需求、持续提升服务质量。人工智能技术的应用效果是影响智慧服务效能的关键。文章基于 TOE 框架与信任理论的核心影响要素,构建了高校图书馆智慧服务效能动态评价的 3 级指标集;创新性采用人机协同方式确定指标权重,即先利用大语言模型模拟评估给出初始权重,人类专家再结合经验进行修正,从而建立高校图书馆智慧服务效能评价体系,并以西北工业大学图书馆智慧咨询服务场景为例,从技术采纳评估和迭代升级评估两个方面验证评价体系的可行性。评价体系融合了技术—组织—环境三个维度的相关指标,实现了高校图书馆智慧服务效能的动态、多维量化评价,可为高校图书馆智慧服务评价工作提供参考。

**关键词** 高校图书馆 智慧服务 TOE 框架 信任理论 动态评价

**分类号** G251

**DOI** 10.16603/j.issn1002-1027.2025.04.002

**引用本文格式** 杨梦洁,侯点可,倚海伦,等.高校图书馆智慧服务效能动态多维评价体系构建与实践[J].大学图书馆学报,2025,43(4):15-23.

## 1 引言

人工智能正在深度重构图书馆的服务业态与知识生态,通过核心技术突破与场景化应用推动高校图书馆向智慧化方向全面转型<sup>[1]</sup>。《教育部高等教育司 2023 年工作要点》中指出“探索推进未来学习中心试点”,要求高校图书馆利用新一代信息技术,打造支撑学习方式变革的新型基层学习组织。《教育部等九部门关于加快推进教育数字化的意见》(教办〔2025〕3 号)进一步指明技术融合路径,强调探索“人工智能+教育”应用场景新范式,推动大模型与教育教学深度融合。基于此,高校图书馆需主动加速人工智能技术在核心业务场景中的深度应用,推动高校图书馆智慧服务转型。

高校图书馆智慧服务正展现出以人工智能技术为驱动、用户信任为基础、馆商协同为支撑的特征。选择更加适用的人工智能技术应用,有助于高校图书馆真正发展为智慧图书馆<sup>[2]</sup>;用户关系是高校图书馆可持续发展的关键,良好且持久的用户关系需以用户信任为基石<sup>[3]</sup>;馆商协同下的技术开发、动态

升级已成为保障图书馆智慧服务持续有效的关键<sup>[4]</sup>。

良好的评价管理机制可协助图书馆制定计划、作出决策、发现问题、优化服务<sup>[5]</sup>。传统的图书馆服务效能评价体系已难以适应当前需求,亟待重构。高校图书馆智慧服务效能评价体系的构建,既要立足系统观念,综合考量技术、组织、环境多维度因素,更要细致关注影响用户信任的要素,才能科学评价图书馆智慧服务效能,确保其持续提升。

## 2 高校图书馆服务效能评价体系研究现状

智慧图书馆是传统图书馆在数智时代的新形态,智慧服务是智慧图书馆的核心<sup>[6]</sup>,因此智慧服务效能评价是传统图书馆服务效能评价的扩展和延伸。

在传统图书馆服务效能评价方面,学者们从不同角度出发,构建了多维度的服务效能评价指标体系。丁冬等从人才培养、教学嵌入、科研嵌入、社会服务和文化遗产创新 5 个维度构建了高校图书馆服

\* 通讯作者:张盛兵,ORCID:0000-0002-2854-729X,邮箱:zhangsb@nwpu.edu.cn。



务效能评价指标体系的基本框架<sup>[7]</sup>;陆红如等从馆内资源与合理配置、经费资助及运行机制、用户服务及用户参与三个方面构建了高校图书馆延伸服务绩效评估体系<sup>[8]</sup>;朱美霖从投入产出的视角,构建了25项具体可量化的评估指标<sup>[9]</sup>;苏瑞竹等从服务准备、服务项目、服务对象三个方面构建了包括21个一级指标和73个二级指标的公共图书馆服务效能评估指标体系雏形<sup>[10]</sup>;孙文娉等基于融合发展概念,对公共图书馆绩效评估中的指标进行了合并和拆分,构建了公共图书馆服务效能评价指标体系<sup>[11]</sup>。由上述研究可知,图书馆服务效能评价需立足其职能展开,既要兼顾投入产出效益,关注用户体验,又要以发展的视角贯穿评价过程。

在智慧图书馆服务效能评价方面,学者主要针对智慧图书馆与传统图书馆评价机理的区别以及指标体系的构建进行了研究。单轫等分析了智慧图书馆与传统图书馆、数字图书馆评价机理的本质区别,提出智慧图书馆评价的三大特征:交互关系的衡量、用户数据的自动化和智能化分析、环境因素的纳入<sup>[12]</sup>;赵磊等基于2011—2021年中国知网收录的文献数据,揭示了我国智慧图书馆评价体系的研究历程,提出评价体系的构建要重视用户体验和服务质量,并注重实证检验<sup>[13]</sup>。以上学者的研究为高校图书馆智慧服务效能评价体系的建立提供了方向性指导。马晓亭针对云图书馆建立了基于读者感知价值和云服务收益的评价指标,并描述了云图书馆服务平台效能评估流程<sup>[14]</sup>;马祥涛基于“全评价”理论,从形式、内容、效用3个维度归类了智慧图书馆的评价对象,并围绕智慧图书馆的根本目的和基本目的设置了相应的评价指标<sup>[15]</sup>,尽管指标体系中涵盖了技术、设施相关指标,但其指标遴选存在一定的主观性,且缺乏对指标权重的设定;金国峰等构建了一套包含智慧馆员、智慧管理、智慧服务、智慧设施4个维度的“ChatGPT+智慧图书馆”建设评价指标体系<sup>[16]</sup>,在智慧服务相关指标中,主要聚焦用户满意度、个性化信息推送与导航以及多语言文化支持等方面,技术层面的指标明显缺失,且在指标权重的确定过程中,表现出与学校本体的强相关性;张颖等从馆员队伍、基础设施、业务管理和用户服务4个维度,构建了智慧图书馆成熟度评价指标体系,并针对首都高校图书馆进行了评估<sup>[17]</sup>,但评价结果的精准性和发展策略的普遍适用性尚存局限。

综上所述,针对智慧图书馆服务效能的评价整体上呈现视角多维、指标多样的特征,现有研究存在着指标遴选主观性强、技术适配性指标缺失、用户信任因素考虑不足等问题,尤其是针对人工智能技术的服务效能评价存在着明显的研究空白,使得传统评价体系难以适配当前技术迭代迅猛、服务场景日趋多元的评价需求。因此,针对高校图书馆智慧服务效能评估体系开展研究,既可填补当前的研究空白,又是智慧图书馆发展的现实需求,具有重要的理论价值和实践意义。

### 3 理论基础

#### 3.1 TOE 框架

技术—组织—环境(Technology—Organization—Environment, TOE)框架<sup>[18]</sup>是一种分析企业采用和实施技术创新的理论模型。企业的采纳行为受到技术、组织、环境多种因素的影响。其中,技术因素包括现有技术及新技术特性、基础设施等,组织因素包括企业规模、结构、文化、管理风格、内部资源等,环境因素包括行业竞争、政策法规、市场需求等。TOE框架强调技术、组织或环境的相互作用和耦合关系,需要三者协同才能达到组织目的<sup>[19]</sup>。

#### 3.2 信任理论

信任是一个关系概念,是人机关系的核心<sup>[20]</sup>。在人工智能领域,信任是委托人在不确定性或脆弱性为特征的情境下,认为代理人能够帮助其实现某个目标的态度<sup>[21]</sup>。信任并非静态存在,而是一个动态的过程<sup>[22]</sup>,会受到人工智能系统性能、用户体验以及外部因素的综合影响。具体来说,人工智能系统性能是人机信任的基石,交互过程中的情感反馈发挥着关键的调节作用,外部环境对用户信任产生了隐性影响。

### 4 高校图书馆智慧服务效能动态多维评价体系的构建

TOE框架有助于高校图书馆理解和分析新技术应用的多维影响因素,为高校图书馆智慧服务效能评价体系提供了理论基础和整体框架。在以人工智能技术驱动的智慧服务过程中,人机信任的发展逻辑与图书馆“以读者为中心”的服务理念高度契合。因此,基于TOE框架和人机信任的关键影响因素遴选相应指标,能够充分体现评价体系的系统性,



突出图书馆核心理念,有效解决当前评价中存在的指标主观性强、技术指标缺失及用户信任因素考量不足等问题。

#### 4.1 评价指标集

##### 4.1.1 技术指标

影响服务效能的常见技术因素主要有响应时间、可扩展性、兼容性、复杂性与可察性<sup>[14, 23]</sup>。从智慧服务特性来看,人机对话作为基础功能,其意图识别能力<sup>[24]</sup>和内容生成可信度<sup>[25]</sup>直接影响服务质量,智慧服务的友好度和同理心对用户满意度起到显著正向影响<sup>[26]</sup>。从用户信任视角分析,影响初始信任的核心因素主要有感知有用性和感知易用性<sup>[27]</sup>,交互过程中,感知信任显著正向影响用户继续使用智能服务的主观意愿,而个人信息泄露、资源质量问题等感知风险则负向影响<sup>[28]</sup>。其中,高校图书馆领域的资源安全问题主要为内容版权风险与意识形态风险。

针对以上因素进行归属分析,技术指标下设置功能特性、复杂程度、感知价值和感知风险 4 个二级指标。其中,功能特性是技术本身的核心能力,包含意图识别能力、响应时间、可信度、可扩展性、情感支持 5 个三级指标;复杂程度包含感知易用性和兼容性 2 个三级指标,分别指向用户端和管理端的技术适配性;感知价值聚焦可感知的技术相对优势,包含感知有用性、服务效率提升、服务质量提升 3 个三级指标;感知风险则重点关注技术引发的潜在风险,设隐私安全风险、内容版权风险、意识形态风险 3 个三级指标,由此实现了对技术层面的多角度评价。

##### 4.1.2 组织指标

组织维度的影响因素主要有管理支持、经费预算、员工水平以及技术培训<sup>[29]</sup>。其中,员工水平和培训同属于人员支持方面,故在组织指标下设置管理层态度、经费预算、人员支持 3 个二级指标。根据当前图书馆智慧服务的馆商协同现状,参与技术服务的人员包括智慧馆员以及研发公司所配备的服务人员。

管理层对新技术采纳的本质取决于管理层对新技术及研发公司的信任程度以及进行技术创新的决心,直接影响管理层的采纳决策和推进力度,据此将管理层态度细分为管理层的信任程度和采纳决心 2 个三级指标;经费预算顺利实施人工智能技术应用的重要保障,具体包括 2 个三级指标;技术的专项经费投入与相应的算力、存储等配套设施预算,其预

算越高,新技术的落地实施难度越大;人员支持方面,主要从数量、能力以及培养角度出发,设置人员数量、知识水平、技能水平以及培训交流 4 个三级指标。经过以上分析,组织指标中包含了 3 个二级指标以及 8 个三级指标,覆盖了组织层面对技术采纳应用提供的全方位支持。

##### 4.1.3 环境指标

根据环境维度中的既有影响因素,在环境指标下设置政策环境、行业竞争力、市场需求 3 个二级指标。政策环境方面,政策监管是保障高校图书馆智慧服务合法合规运行的基础,政策引导体现国家层面对新技术的推广导向;行业竞争力方面,主要有两方面内容:一是技术层面的竞争,新技术与社会发展趋势的契合度越高,越能体现其先进性;二是图书馆的行业竞争,新技术在其他高校中采纳越多,其竞争力越强,越能推动采纳决策;市场需求层面,主要指高校用户的需求,技术功能与用户需求匹配度越好,越有利于其落地。

据此,环境指标下设置政策环境、行业竞争力和市场需求 3 个二级指标。其中,政策环境包含政策监管、政策引导 2 个三级指标,行业竞争力包含趋势契合度、同行采纳 2 个三级指标,市场需求方面包含了需求满足度 1 个三级指标。

综合归纳起来,高校图书馆智慧服务效能评价指标集共包含 3 个一级指标、10 个二级指标、26 个三级指标,相应的指标解释详见表 1。

#### 4.2 指标权重确定

指标权重在评价体系中至关重要,它反映了指标间的相对重要性。大模型因其强大的知识储备、整合和分析能力,能够从多维度进行感知、推理和决策<sup>[30]</sup>,可利用大模型和人类专家协作评估的方式,有效提高权重确定的准确性和效率。

##### 4.2.1 大模型模拟赋值

选取中国知网 AI 平台、玻尔科研空间站、AMiner 三个基于大量的科研文献资源训练的大模型产品作为模拟专家,针对高校图书馆智慧服务效能评价指标集进行权重赋值,取三者的平均值为模拟权重。

三个大模型对于一级指标的相对重要性反映出相同的认知,认为技术指标是评估高校图书馆智慧服务效能的核心,即技术指标权重最高,为 40%;组织指标次之,平均权重为 32%;环境指标平均权重为 28%,相对最低。



表 1 高校图书馆智慧服务效能评价指标集

| 一级指标          | 二级指标           | 三级指标           | 指标解释                               |                          |
|---------------|----------------|----------------|------------------------------------|--------------------------|
| 技术指标<br>(TF)  | 功能特性<br>(TF1)  | 意图识别能力(TF11)   | 是否准确识别用户询问、请求或指示的目的                |                          |
|               |                | 响应时间(TF12)     | 从问到答所需的时间                          |                          |
|               |                | 可扩展性(TF13)     | 是否支持模型替换或升级、功能及知识库扩展、可嵌入或被嵌入其他业务系统 |                          |
|               |                | 可信度(TF14)      | 数据来源、执行策略、输出结果是否可信,是人机信任的重要影响因素    |                          |
|               |                | 情感支持(TF15)     | 根据用户的情绪状态和需求,提供个性化关怀和建议            |                          |
|               | 复杂程度<br>(TF2)  | 感知易用性(TF21)    | 访问界面、使用方式是否简单方便,直接影响人机信任           |                          |
|               |                | 兼容性(TF22)      | 与现有软硬件的兼容程度                        |                          |
|               | 感知价值<br>(TF3)  | 感知有用性(TF31)    | 用户主观上认为技术有用的程度,直接影响人机信任            |                          |
|               |                | 服务效率提升(TF32)   | 可感知的服务效率提升程度                       |                          |
|               |                | 服务质量提升(TF33)   | 可感知的服务质量提升程度                       |                          |
|               | 感知风险<br>(TF4)  | 隐私安全风险(TF41)   | 是否存在泄露个人数据的风险                      |                          |
|               |                | 内容版权风险(TF42)   | 数据来源是否存在版权风险                       |                          |
|               |                | 意识形态风险(TF43)   | 数据来源、输出结果是否有意识形态风险                 |                          |
|               | 组织指标<br>(OF)   | 管理层态度<br>(OF1) | 信任程度(OF11)                         | 管理层对新技术、研发公司的信任程度        |
|               |                |                | 采纳决心(OF12)                         | 管理层对新技术应用的决心是否强烈         |
|               |                | 经费预算<br>(OF2)  | 专项经费预算(OF21)                       | 新技术的经费预算额度,预算越高,越不利于技术应用 |
| 配套经费预算(OF22)  |                |                | 所需的配套设施经费预算额度,如算力、存储等              |                          |
| 人员支持<br>(OF3) |                | 人员数量(OF31)     | 支持智慧服务的人员数量(含馆员、研发公司人员)            |                          |
|               |                | 知识水平(OF32)     | 人员的受教育程度                           |                          |
|               |                | 技能水平(OF33)     | 人员的专业技能水平                          |                          |
|               |                | 培训交流(OF34)     | 是否提供相关的培训交流                        |                          |
| 环境指标<br>(EF)  | 政策环境<br>(EF1)  | 政策监管(EF11)     | 研发公司及技术应用是否符合国家相关监管政策              |                          |
|               |                | 政策引导(EF12)     | 新技术应用是否符合国家、机构等的政策引导               |                          |
|               | 行业竞争力<br>(EF2) | 趋势契合度(EF21)    | 新技术是否紧跟社会发展趋势,保持先进性                |                          |
|               |                | 同行采纳(EF23)     | 新技术应用在其他高校中的采纳情况,采纳数量越多,竞争力越强      |                          |
|               | 市场需求<br>(EF3)  | 需求满足度(EF31)    | 新技术满足用户需求的程度                       |                          |

二级指标中,功能特性以 16% 的权重居首,凸显其在整体评价体系中的核心地位;感知价值、管理层态度与政策环境的模拟权重均为 11%;经费预算、人员支持与行业竞争力的模拟权重均为 10%;

复杂程度和市场需求分别以 9% 和 8% 的平均权重位列其后,重要性相对弱于前几类指标;大模型认为感知风险的相对重要性最低,平均权重仅为 4%。

三级指标模拟权重的梯度差异更为细致。市场需



求指标下的需求满足度以 8% 的权重位居前列,体现了大模型对需求导向的高度认可;感知易用性、信任程度、专项经费预算、政策监管指标的模拟权重均为 6%,显示大模型对用户体验与组织资源保障、政策监管的综合关注度;意图识别能力、感知有用性、政策引导、趋势契合度、同行采纳五项指标的模拟权重为 5%;响应时间的模拟权重 4% 低于上述因素,暗示大模型的评价逻辑中对效果的重视程度高于速度;模拟权重为 3% 的指标数量最多,涵盖可扩展性、可信度、兼容性、服务效率提升、服务质量提升、人员数量、知识水平 7 项,处于辅助性地位;隐私安全风险、技能水平、培训交流以 2% 的模拟权重位列其后,反映大模型对潜在风险与长期能力的关注度不足;情感支持、内容版权风险、意识形态风险的权重最低,仅为 1%,与二级指标中感知风险的弱势地位形成呼应,凸显了大模型在风险维度考量上的显著局限性。

#### 4.2.2 人类专家修正权重

针对大模型模拟赋权的局限性,研究遴选了 6 名人类专家参与权重的修正。人类专家主要由两

类人员组成:3 名高校图书馆馆长以及 3 名从事人工智能技术工作的专业人员。人类专家基于自身专业知识和实践经验,对大模型赋值结果进行调整。

人类专家认为一级指标中技术指标的模拟权重最高,符合基本观点,印证了智慧服务是由技术驱动的建设规律。二级指标中,市场需求和感知风险的模拟权重偏低,与高校图书馆为用户提供高质量信息资源的核心服务职能、教育职能不匹配;管理层态度、经费支持、人员支持、政策环境、行业竞争力的权重偏高。为了进一步强化高校图书馆智慧服务的用户需求导向,凸显图书馆作为可信资源提供者的特点,人类专家重点针对三级指标的模拟权重进行调整,调整结果详见表 2。

经过调整,二级指标中的功能特性权重提升 1%,感知风险和用户需求权重分别大幅提升至 10% 和 15%;管理层态度、经费预算、人员支持权重下降 2%,政策环境和行业竞争力权重下降 3%。相应的一级指标权重调整为:技术指标 45%,组织指标 25%,环境指标 30%。

表 2 高校图书馆智慧服务效能评价指标模拟权重和调整权重对比

| 一级指标         | 模拟权重 | 调整权重 | 二级指标          | 模拟权重 | 调整权重 | 三级指标          | 模拟权重 | 调整权重 | 增减情况         |
|--------------|------|------|---------------|------|------|---------------|------|------|--------------|
| 技术指标<br>(TF) | 40%  | 45%  | 功能特性<br>(TF1) | 16%  | 17%  | 意图识别能力(TF11)  | 5%   | 5%   |              |
|              |      |      |               |      |      | 响应时间(TF12)    | 4%   | 4%   |              |
|              |      |      |               |      |      | 可扩展性(TF13)    | 3%   | 3%   |              |
|              |      |      |               |      |      | 可信度(TF14)     | 3%   | 3%   |              |
|              |      |      |               |      |      | 情感支持(TF15)    | 1%   | 2%   | +            |
|              |      |      | 复杂程度<br>(TF2) | 9%   | 6%   | 感知易用性(TF21)   | 6%   | 4%   | -            |
|              |      |      |               |      |      | 兼容性(TF22)     | 3%   | 2%   | -            |
|              |      |      |               |      |      | 感知有用性(TF31)   | 5%   | 4%   | -            |
|              |      |      | 感知价值<br>(TF3) | 11%  | 12%  | 服务效率提升(TF32)  | 3%   | 4%   | +            |
|              |      |      |               |      |      | 服务质量提升(TF33)  | 3%   | 4%   | +            |
|              |      |      |               |      |      | 感知风险<br>(TF4) | 4%   | 10%  | 隐私安全风险(TF41) |
|              |      |      | 内容版权风险(TF42)  | 1%   | 3%   |               |      |      | +            |
|              |      |      | 意识形态风险(TF43)  | 1%   | 3%   |               |      |      | +            |



续表

| 一级指标         | 模拟权重 | 调整权重 | 二级指标           | 模拟权重 | 调整权重 | 三级指标         | 模拟权重 | 调整权重 | 增减情况 |
|--------------|------|------|----------------|------|------|--------------|------|------|------|
| 组织指标<br>(OF) | 32%  | 25%  | 管理层态度<br>(OF1) | 11%  | 9%   | 信任程度(OF11)   | 6%   | 5%   | -    |
|              |      |      |                |      |      | 采纳决心(OF12)   | 5%   | 4%   | -    |
|              |      |      | 经费预算<br>(OF2)  | 10%  | 8%   | 专项经费预算(OF21) | 6%   | 5%   | -    |
|              |      |      |                |      |      | 配套经费预算(OF22) | 4%   | 3%   | -    |
|              |      |      |                |      |      | 人员数量(OF31)   | 3%   | 2%   | -    |
|              |      |      | 人员支持<br>(OF3)  | 10%  | 8%   | 知识水平(OF32)   | 3%   | 2%   | -    |
|              |      |      |                |      |      | 技能水平(OF33)   | 2%   | 2%   | -    |
|              |      |      | 培训交流(OF34)     | 2%   | 2%   | -            |      |      |      |
| 环境指标<br>(EF) | 28%  | 30%  | 政策环境<br>(EF1)  | 11%  | 8%   | 政策监管(EF11)   | 6%   | 5%   | -    |
|              |      |      |                |      |      | 政策引导(EF12)   | 5%   | 3%   | -    |
|              |      |      | 行业竞争力<br>(EF2) | 10%  | 7%   | 趋势契合度(EF21)  | 5%   | 3%   | -    |
|              |      |      |                |      |      | 同行采纳(EF22)   | 5%   | 4%   | -    |
|              |      |      | 市场需求<br>(EF3)  | 8%   | 15%  | 需求满足度(EF31)  | 8%   | 15%  | +    |

## 5 实践应用——以西北工业大学图书馆智慧咨询服务场景为例

### 5.1 技术采纳评估

面对 AI 技术浪潮,西北工业大学图书馆管理层展现出高度的战略前瞻性与积极行动力,成立了由馆长牵头、技术骨干组成的人工智能应用研究小组,统筹智慧服务的规划与落地,并对小组成员提供培训支持和交流,持续提升团队的技术认知与能力。

通过广泛调研师生智慧服务需求,小组明确了现阶段的用户需求主要有知识问答、资源发现、馆务咨询等。在此基础上,围绕智慧服务效能评价体系设计测试问题,并对其进行意图解析,明确考察指标;进而制定评分细则,建立可量化的评价标准;依托评价细则,对拟采纳的两种智慧咨询应用开展测试,测试结果如图 1 所示。

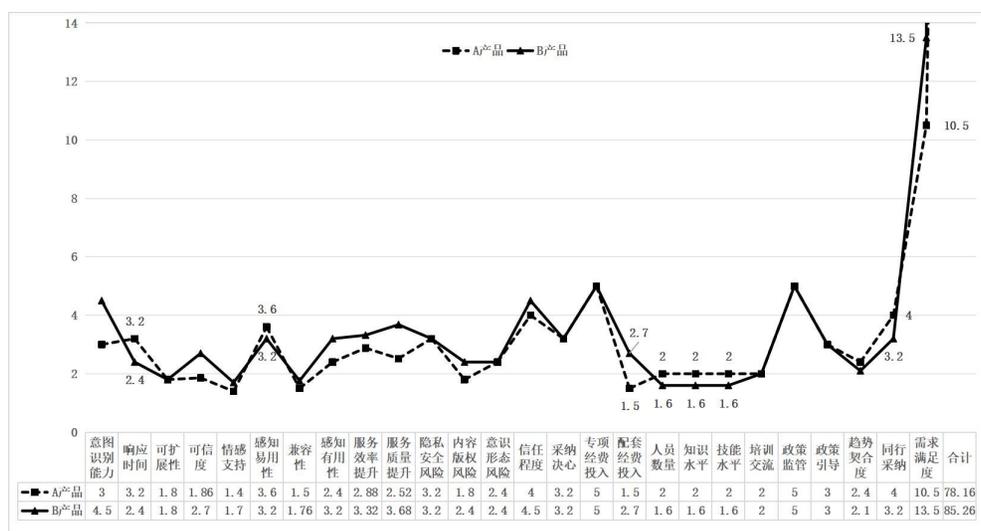


图 1 A 产品与 B 产品测评打分情况



从图 1 中可以看出:A 产品在响应速度、人员支持和同行采纳上具有优势;B 产品凭借更高的意图识别能力,在用户需求满足度和感知价值方面表现突出,且与现有系统的兼容性更佳,所需的配套预算也相对较低。综合来看,B 产品展现出更优的综合性能。

### 5.2 技术迭代升级评估

针对 B 产品各项指标得分进行分析,发现其响应时间、可扩展性及趋势契合度三项指标得分偏低。小组聚焦这三项指标开展重点分析,梳理关键问题

后与研发公司联合研讨改进方向,最终制定针对性改进方案(见表 3)。

鉴于高校图书馆智慧服务效能评价体系是技术—组织—环境的动态适配过程,小组依托该指标体系对升级后的智慧服务效能进行复评,升级前后各指标的归一化得分对比情况如图 2 所示。从图 2 中可以看出:通过对响应时间与可扩展性的改进,促进了多项技术指标的提升,改善了用户需求满足度,但同时增加了配套设施预算及隐私安全风险,导致相应指标的得分略有下降。

表 3 B 产品迭代升级方案

| 改进指标  | 存在问题                    | 改进建议  |
|-------|-------------------------|---|
| 响应时间  | 受任务类型、用户需求及硬件算力多重影响     | <ol style="list-style-type: none"> <li>1. 优先判断馆务咨询、信息查询等常见类型,快速响应常见需求</li> <li>2. 实施智能的缓存策略,基于历史查询和热门问题进行预加载,减少重复计算,缩短响应时间</li> <li>3. 部署高性能硬件</li> </ol> |
| 可扩展性  | 仅支持 2 种模型,选择受限,对接系统数量偏少 | <ol style="list-style-type: none"> <li>1. 支持多模型并行部署</li> <li>2. 提供标准化 API,与校园信息生态融合</li> <li>3. 建立知识库自动化采集机制</li> </ol>                                   |
| 趋势契合度 | 对行业技术发展响应不够灵敏           | <ol style="list-style-type: none"> <li>1. 设立专职团队跟踪 AI 技术趋势</li> <li>2. 常态化市场调研,调整产品方向</li> <li>3. 分析竞争对手策略,制定差异化方案</li> </ol>                             |

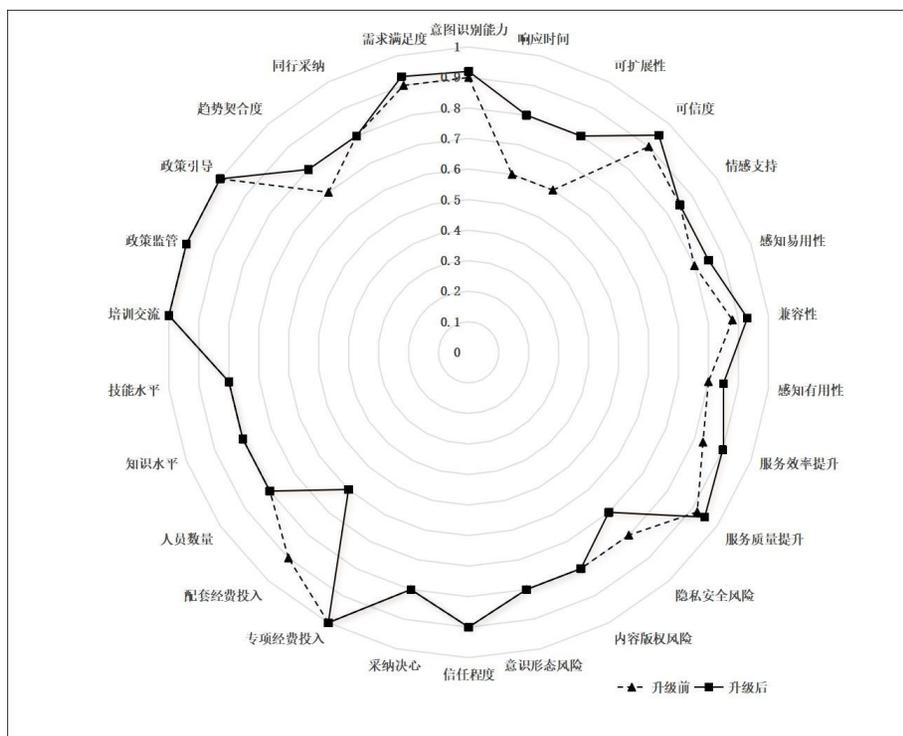


图 2 B 产品升级前后测评对比情况



## 6 结论与建议

本文基于 TOE 框架的多维耦合因素与人机信任理论的核心影响因素,分析确定了影响高校图书馆智慧服务效能的 3 级指标集,并依托人机协同机制,发挥大语言模型与人类专家的各自优势,高效率确定了指标权重,成功构建起高校图书馆智慧服务效能的动态量化评估体系。以西北工业大学图书馆智慧咨询服务为例,验证了该体系的适用性与动态调整能力,实现了技术的采纳评估和迭代升级评估,推动了服务效能提升。

在高校图书馆智慧服务效能评价体系的推广应用方面,应关注以下四点:

### (1) 强化多维评价思维,推动管理变革

推动技术落地过程中,需结合组织自身特性、环境条件进行多维度评价,确保技术选型与落地始终依据自身实际、聚焦用户需求。同步优化管理机制,推动管理团队从技术驱动向效能驱动转变,实现“以评促建”“以评促效”的闭环管理。

### (2) 分类调整评价指标,促进多场景适用

人工智能技术应用中,不同场景的目标、约束与风险差异显著,高校图书馆智慧服务效能评价体系需结合场景特征、技术发展、运行数据反馈调整指标集及权重,方能更加贴合不同场景需求,为技术合理应用与持续改进提供支撑。

### (3) 建立动态评价机制,驱动服务效能提升

建立动态化指标数据监测机制,跟踪用户行为数据、需求变化、技术发展态势、政策环境变化等,定期开展服务效能评价。依据动态数据分析和评价结果,改进技术应用,优化相应指标,促进技术升级和服务效能提升。

### (4) 探索人机协同评估模式,提升评估效率

大模型在评估领域正展现出突破性应用潜力,其技术特性为破解传统评估模式的痛点提供了全新路径。大模型与专家经验的协同模式可实现技术赋能和专家把关的双重价值,大模型通过持续吸收专家修正意见与新场景数据,可逐步提升对复杂评估场景的适配性,推动评估工作向更精准、高效、智能的方向演进。

## 参考文献

1 陈燕方,储继华,刘后滨. 高校图书馆服务转型的时代背景与创新发展的历史任务[J]. 大学图书馆学报, 2024, 42(2): 32-37.

2 王毅,董怡婷. 类 ChatGPT 人工智能在图书馆智慧服务中的应用与思考[J]. 图书馆理论与实践, 2023(6): 129-136.

3 陈建龙,赵飞,及桐. 基于信任理论的北京大学图书馆用户关系管理新探[J]. 大学图书馆学报, 2024, 42(1): 11-16.

4 张小燕,雷苗. 人工智能场景下图书馆服务面临的挑战及应对策略[J]. 图书与情报, 2024(6): 113-121.

5 吴建华. 数字图书馆评价方法[M]. 北京: 科学出版社, 2009: 24.

6 初景利,段美珍. 智慧图书馆与智慧服务[J]. 图书馆建设, 2018(4): 85-90.

7 丁冬,张长秀. 高校图书馆服务效能评价指标体系框架初探[J]. 图书馆工作与研究, 2018(10): 58-63.

8 陆红如,陈雅. 基于服务效能提升的高校图书馆延伸服务绩效评估体系构建[J]. 大学图书馆学报, 2019, 37(2): 17-23.

9 朱美霖. 投入产出视角下的公共图书馆文化服务效能评估模型研究[J]. 图书馆研究与工作, 2021(2): 14-21.

10 苏瑞竹,李梦菲. 利益相关者视角下公共图书馆服务效能评估指标体系框架初探[J]. 图书馆理论与实践, 2022(6): 38-47.

11 孙文婷,陈雅. 融合发展语境下公共图书馆服务效能评价体系研究[J]. 图书馆建设, 2023(6): 155-167.

12 单珍,陈雅. 我国智慧图书馆评价理论探析[J]. 图书馆学研究, 2022(9): 2-9.

13 赵磊,余吕娜,高滢,等. 我国智慧图书馆评价体系研究可视化分析[J]. 图书馆工作与研究, 2023(4): 65-74.

14 马晓亭. 复杂云计算环境下基于客户感知价值的数字图书馆服务效能评估[J]. 图书馆理论与实践, 2014(3): 84-86.

15 马祥涛. “全评价”视角下智慧图书馆建设评价研究[J]. 图书馆理论与实践, 2023(3): 103-109.

16 金国峰,杨毅豪,王圆圆. “ChatGPT+智慧图书馆”建设评价指标体系构建[J]. 新世纪图书馆, 2025(3): 15-23.

17 张颖,段美珍,刘惠,等. 首都高校智慧图书馆建设成熟度评价与发展策略研究[J]. 图书情报工作, 2025, 69(5): 28-38.

18 Baker J. The technology-organization-environment framework [M]//Dwivedi Y K, Wade M R, Schneberger S L. Information Systems Theory: Explaining and Predicting Our Digital Society. New York, NY: Springer New York, 2012: 231-245.

19 罗杨洋,周国辉,韩锡斌. 高校数字化转型如何适配有效策略?——基于技术、组织、环境协同的视角[J]. 现代教育技术, 2025, 35(6): 46-55.

20 张洪忠,任吴炯. 超越“第二自我”的人机对话——基于 AI 大模型应用的信任关系探讨[J]. 新闻大学, 2024(3): 47-60.

21 吴俊,张迪,刘涛,等. 人类对人工智能信任的接受度及脑认知机制研究:实证研究与神经科学实验的元分析[J]. 管理工程学报, 2024, 38(1): 60-73.

22 Virvou M, Tsihrintzis G A, Tsihrintzi E. VIRTISI: a novel trust dynamics model enhancing artificial intelligence collaboration with human users-insights from a ChatGPT evaluation study[J]. information Sciences, 2024, 675: 120759.

23 刘细文,金学慧. 基于 TOE 框架的企业竞争情报系统采纳影响因素研究[J]. 图书情报工作, 2011, 55(6): 70-73.

24 刘娇,李艳玲,林民. 人机对话系统中意图识别方法综述[J]. 计算机工程与应用, 2019, 55(12): 1-7.

25 钱明辉,李胡蓉,杨建梁. 大语言模型可信:内涵、影响、挑战与对策[J]. 图书情报工作, 2024, 68(20): 69-86.



- 26 王晰巍, 罗然, 刘宇桐, 等. 智慧图书馆在线聊天机器人使用行为影响因素及实证研究[J]. 情报学报, 2023, 42(2): 217-230.
- 27 冉龙亚, 陈涛, 孙宁华. 技术特性对政务智能聊天机器人初始信任的影响[J]. 情报杂志, 2022, 41(7): 138-144.
- 28 范昊, 徐颖慧, 曾子明. 智慧图书馆 AI 服务用户接受行为影响因素研究[J]. 图书馆学研究, 2021(2): 37-47.
- 29 Al-Khatib A W. Drivers of generative artificial intelligence to fostering exploitative and exploratory innovation: a TOE framework[J]. Technology in Society, 2023, 75: 102403.
- 30 Gao C, Lan X, Li N, et al. Large language models empowered agent-based modeling and simulation: a survey and perspectives [J]. Humanities and Social Sciences Communications, 2024, 11: 1259.

作者贡献说明:

杨梦洁: 选题调研、研究思路构想, 论文撰写及修改

侯点可: 研究内容讨论, 数据汇总

倚海伦: 研究内容讨论, 数据分析

黄辉: 研究内容讨论, 论文修改

张盛兵: 论文选题、论文结构、论文修改

作者单位: 西北工业大学图书馆, 陕西西安, 710072

收稿日期: 2025 年 7 月 10 日

修回日期: 2025 年 7 月 17 日

(责任编辑: 关志英)

## Construction and Practice of Dynamic Multi-dimensional Evaluation System for Smart Service Effectiveness in University Libraries

YANG Mengjie HOU Dianke YI Hailun HUANG Hui ZHANG Shengbing

**Abstract:** The evaluation of service effectiveness in university libraries is of great significance as it aims to guide libraries to continuously meet users' needs and steadily improve service quality. The application of Artificial Intelligence (AI) technology serves as a key factor affecting the effectiveness of smart services. Traditional evaluation systems have inherent limitations, including excessive subjectivity in metric selection, a critical lack of technical performance indicators, and insufficient consideration of users' trust in AI-driven services. To address these limitations, this paper develops a three-level metric set for dynamically evaluating the smart service effectiveness of university libraries, drawing on the core influencing factors of the TOE (Technology, Organization, Environment) framework and trust theory. This set includes 3 first-level indicators, 10 second-level indicators, and 26 third-level indicators. This metric set not only incorporates the multi-dimensional coupling factors at the technical, organizational, and environmental levels for smart service effectiveness evaluation, but also embodies the library's "reader-centered" service philosophy. We innovatively adopted a human-machine collaboration method to determine the index weights. First, we leveraged the large language model's robust knowledge reserve, integration, and analysis capabilities to simulate evaluations and generate initial weights for the metrics. Then, we consulted human experts to revise these initial weights based on their professional experience. This method effectively improved the accuracy and effectiveness of weight determination. The metric set and corresponding weights together constituted the evaluation system for the smart service effectiveness of university libraries. Taking the smart consulting service scenario of the Northwestern Polytechnical University Library as an example, we verified the applicability and dynamics of the evaluation system from two aspects: technology adoption assessment and iterative upgrading assessment. The evaluation system integrated relevant metrics from the three dimensions of technology, organization, and environment, filling the research gap in the evaluation of the service effectiveness of artificial intelligence technology in university libraries. It realized dynamic, multi-dimensional, and quantitative evaluation of the smart service effectiveness in university libraries, thus boasting significant theoretical value and practical significance. Finally, the paper put forward the following suggestions for the smart service evaluation of university libraries: strengthening multi-dimensional evaluation thinking to promote management reform, adjusting evaluation indexes by category to facilitate application in multiple scenarios, establishing a dynamic evaluation mechanism to drive the improvement of service effectiveness, and exploring a human-machine collaborative evaluation model to enhance evaluation effectiveness.

**Keywords:** University Library; Smart Service; TOE Framework; Trust Theory; Dynamic Evaluation