

大语言模型对我国公共图书馆价值的感知差异研究*

陈慧彤 郑尔特 闫 慧 赵雪锋

摘要 探究大语言模型(LLM)相较于人类主体而言对我国公共图书馆价值的感知差异,可以为生成式人工智能背景下提升公共图书馆的社会价值提供参考。基于文献研究和我国35所公共图书馆的品牌活动案例设计问题,分别在GPT-4、文心一言、DeepSeek等LLM上进行测试,并以馆员和用户的测试结果作为基准,分析LLM对于公共图书馆六类服务价值的感知水平及差异。研究发现,LLM、馆员和用户都对公共图书馆的“社会教育”“文化服务”“阅读服务”给予高度认可,“健康服务”则被三者赋予最低的评价;不同LLM之间存在感知差异,国内LLM对公共图书馆的多类服务持较高评价,而国外LLM则持相对较低评价;LLM对公共图书馆价值的认可度低于人类,尤其是在“信息服务”领域。图4。表3。参考文献55。

关键词 生成式人工智能 大语言模型 公共图书馆价值 感知差异 感知价值

Research on the Perception Differences of Large Language Models Regarding the Value of Public Libraries in China

Chen Huitong Zheng Er-te Yan Hui Zhao Xuefeng

Abstract: The article aims to explore the perception differences of the value of public libraries in China between Large Language Models (LLMs) and humans, providing a reference for enhancing the social value of public libraries in the context of AI-Generated Content (AIGC). Based on literature research and brand activity cases from 35 public libraries in China, questions were designed and tested on GPT-4, Ernie Bot, and DeepSeek. The test results from librarians and users were used as benchmarks to analyze and evaluate the LLMs' level of perception and differences regarding six types of service values provided by public libraries in detail. The study found that LLMs, librarians, and users all highly recognize the roles of public libraries in “social education”, “cultural services”, and “reading services”, while “health services” are given the lowest evaluation by all three. There are cognitive differences among different LLMs. Domestic LLMs tend to have a higher evaluation of most service types, while foreign LLMs exhibit a relatively lower evaluation. The recognition of the value of public libraries by LLMs is lower than that of humans, especially in the field of information services. 4 figs. 3 tabs. 55 refs.

Keywords: Generative AI; Large Language Model; Public Library Value; Perception Difference; Perceived Value

长期以来,公共图书馆都面临着价值感知和社会影响力的问题。2011年杭州图书馆的“乞丐入馆”、2020年的“农民工读者留言东莞图书馆”,这两个21世纪以来中国公共图书馆最为轰动的新闻热点,在给公众带来温情与感动的同时,也

让公共图书馆的功能和价值得到了广泛的传播。但这也引发了对图书馆信息职业的反思:为什么这些看似公共图书馆最为基础、常规的业务,会成为引发公众热议的爆点?可见,公共图书馆的核心价值和社会作用仍未深入人心,图书馆的社会

* 本文系国家社会科学基金重点项目“人工智能对国民经济的颠覆性影响及信息治理研究”(项目编号:23AZD093)研究成果之一。

通讯作者:郑尔特,电子邮箱:ezheng1@sheffield.ac.uk。

认知度存在非常大的提升空间^[1]。

当前,公共图书馆的服务营销和推广越来越多地以用户为中心,从用户视角去评估和改进服务质量^[2]。LibQUAL+™ 模型为感知质量评价提供了一个系统的工具,被广泛应用于图书馆的服务质量评估^[3]。但随着生成式人工智能(AIGC)的快速发展和应用,用户视角的图书馆服务评估工作可能会面临一定的挑战。2024年7月,OpenAI正式进军搜索领域,发布了SearchGPT原型^[4],可以预见的是,未来将有大量的搜索引擎、服务应用、生产工具会内嵌ChatGPT等大语言模型(Large Language Model, LLM)。目前已经有很多研究关注到了LLM对于用户决策的影响以及人类和LLM的协同决策问题,LLM可以通过分析已有数据来辅助人类决策^[5],也能够更快、更低成本地生成创意方案^[6],并对已有决策进行优化^[7]。鉴于LLM对用户的影响,有必要在用户中心视角的基础上,关注LLM对我国公共图书馆价值的感知水平,探究LLM相较于人类而言对公共图书馆价值的感知差异,这对于理解和提升公共图书馆服务价值、普及和发展公共图书馆事业具有重要意义。

1 文献综述

如何证明公共图书馆的价值,一直是图书馆界关注的议题。由于图书馆难以直接计算产出与投入,因此需要根据用户感知来评估图书馆价值。其中,基于投资回报率和成本效益分析的经济价值评估^[8]一直是研究的重点,图书馆的价值被定义为感知效益与感知成本的比值^[9]。条件价值评估法(CVM)的应用最为广泛,通过调查公众为获得图书馆服务的支付意愿(WTP)或剥夺公众获得图书馆服务的权利的接受补偿意愿(WTA),来量化公共图书馆的经济价值^[10-12]。在此基础上,学者制定了完善的经济价值指标体系,对图书馆员和用户的图书馆经济价值感知进行测量和比较^[13]。除此之外,公共图书馆的社会价值^[14]、“场所”价值^[15]、教育价值^[16]、健康服务价值^[17]也被

研究所证实。相关研究通过比较图书馆员和用户的价值感知发现,图书馆员和用户群体对图书馆的教育价值评分均值最高,社会价值评分次之,对经济价值的评价则最低,大多数图书馆员对图书馆的评价比用户群体更高^[18]。进一步分析图书馆社会价值的感知差异,大多数图书馆员比用户更倾向于平等地评估所有指标,用户则对图书馆提升弱势群体社会适应力的价值更为认同^[19]。

与价值密切关联的概念是影响力。影响力通常与产出或成效有关^[19],指图书馆及其服务对个人或社会产生的影响和改变^[20]。因此,图书馆服务的价值能否在用户实际使用中体现,是影响力评价的主要关注点^[20]。关于影响力评价的指标体系和方法流程已经有了丰富的积累^[21,22],评估的支持性证据包括馆员或者用户的感知数据^[20,23,24]、第三方媒体报道^[25,26]、政策文本、经典案例、历史文献^[27]等,评估的方法强调利用不同来源和类型的数据、基于追踪调查的变化理论视角^[28],来增强评价结果的精确性和说服力。除了影响力评价,公众对公共图书馆的积极形象感知也是证明图书馆价值的有效方式。公共图书馆的社会形象被定义为公众心目中关于公共图书馆的相对稳定的地位和整体映像^[29]。研究视角包括:(1)公共图书馆供给方维度,包括建筑形象^[30]、馆员形象^[31]、文化品牌形象^[32]。(2)图书馆用户主体维度,通过分析公众在小红书^[33]、旅游网站^[34]等平台发布的评论,探索公众对公共图书馆的形象感知。(3)媒介报道维度,关注公共图书馆在不同媒介中的形象表现,从报纸^[35]、科幻小说^[36]、网络新闻^[37]等内容中提取和总结公共图书馆的形象描述。

综上所述,公共图书馆的价值感知研究已经覆盖了经济、社会、教育等多个主题领域,也使用了服务供给方、需求方、第三方等多个来源的评价数据。但是,在LLM在各应用场景中逐步普及的当下,LLM对公共图书馆价值的感知问题仍未得到充分的研究。目前有不少研究试图揭示LLM在模拟人类认知过程方面的潜力^[38-40],特别是在理解和解释复杂的社会结构和交互方面^[41]。这

方面的研究在心理学、社会学和政治学等众多计算社会科学领域已经取得了一定的进展。其中,LLM 存在的偏见问题也逐渐浮出水面,性别偏见^[42]、地理偏见^[43]、政治偏见^[44]等问题引起了众多学者的关注和讨论,但鲜有研究对 LLM 的公共图书馆价值感知差异进行测量。

基于此,本研究将 LLM 对公共图书馆价值的感知定义为:LLM 通过学习互联网中存在的海量文本数据,形成的对图书馆价值的直接体验和内在表征。本研究的具体问题是:

(1) LLM 对我国公共图书馆价值的感知情况如何?

(2) 以公共图书馆馆员和用户的价值感知作为基准,LLM 对我国公共图书馆的价值感知是否与之存在差异? 差异的表现和影响是什么?

2 研究设计

2.1 问题设计

为了更好地比较 LLM、公共图书馆馆员和用户对公共图书馆价值的感知水平,本研究采用统一的问题设计对三者进行询问,以便在数据分析阶段能够转换为统一的尺度进行计算。

问题需要尽可能地全面涵盖公共图书馆的

价值,因此本研究的问题设计广泛参考了公共图书馆的相关文献研究和实践工作内容。文献研究涉及图书馆学领域的行业标准指南^[45,46]、学者对公共图书馆功能、价值的阐述^[47-49],这方面代表了理想层面上的公共图书馆价值。实践内容则来自于我国公共图书馆的实践案例,本研究调研了 35 所公共图书馆的官方网站和微信公众号,并结合对馆员的访谈,最终筛选得到 182 个品牌活动案例,这方面代表现实层面上的公共图书馆价值。基于两方面的调研,将公共图书馆的价值提炼为 6 个类型,分别为“阅读服务”“信息服务”“数字服务”“社会教育”“文化服务”“健康服务”。每类服务下设计了 5 个问题,最终共设计出 30 个问题,示例见表 1。

30 道题目的选项设计均采用李克特五级量表的形式,选项从低到高依次为:完全没做到(1分)、基本没做到(2分)、效果可有可无(3分)、基本做到(4分)、做的效果非常好(5分)。为了避免填写者不了解问题情境是否属于公共图书馆职责范围而导致的感知差异,本研究增加了“和公共图书馆无关(0分)”的独立选项。在实际问卷设计中,不同类型的问题被随机穿插排序,以免填写者在答题时产生思维惯性。

表 1 公共图书馆价值感知测试题示例

价值类型	问题描述
阅读服务	“5 分钟读完一本名著”“3 分钟看完一部电影”“10 分钟带你了解历史”……人们越来越喜欢快餐文化。培养公众良好的阅读习惯、深度思考能力显得非常迫切。您认为图书馆发挥作用了吗?
信息服务	在反诈骗宣传、增强群众的防骗意识和识骗能力方面,您认为图书馆发挥作用了吗?
社会教育	弘扬志愿服务精神,鼓励更多人为社会出一份力,在这方面您认为图书馆发挥作用了吗?
健康服务	糖尿病患者想要学习一些关于降低血糖的营养饮食等健康信息,在提供这些信息方面,您认为图书馆发挥作用了吗?
数字服务	人工智能越来越火,我们得学会利用它。在帮助公众学习什么是人工智能里的算法、怎样更好地应对算法方面,您认为图书馆发挥作用了吗?
文化服务	在古镇的改建中,只有深入研究地方志、地方史料、家谱等,才能更好地保护其文化内涵。在这些资源的查阅和获取方面,您认为图书馆发挥作用了吗?

在实际问卷设计中,考虑到 LLM 在默认情况下可能产生多样且不可预测的输出,答案具有一定的偶然性,因此,本研究在提示中要求 LLM 在给出评分选项之后,须回答相应理由。针对馆员和用户的问卷中插入了陷阱题(要求填写者必须选择某一选项,若未选该选项则视为无效问卷),以便数据分析阶段筛选无效问卷。问卷的第二部分为个人信息题,针对馆员的问卷题项包括图书馆所属区域和级别、部门岗位及职称等级、在图书馆的工作时间等;针对用户的问卷题项为性别、年龄、教育背景、职业、常住地区,以及使用公共图书馆的频率等。

2.2 基准数据收集

在预调研阶段,本研究邀请了 5 位公共图书馆馆员和 10 位公共图书馆用户填写问卷,据其意见对问卷表述进行了优化,使跨年龄段的填写者均能更好地理解问题。2024 年 7 月 20 日至 31 日,本研究采用线上方式发放正式问卷。馆员问卷共发放 60 份,剔除无效问卷后得到有效问卷 54 份,样本涵盖了我国东部、中部和西部的省级、市级和县(区)级公共图书馆。图书馆用户问卷共发放 160 份,剔除无效问卷后得到有效问卷 147 份。其中,男性有 61 人(占比 41.5%),女性有 86 人(占比 58.5%);各个年龄段均有所覆盖,40.5%的调研对象年龄在 18—25 岁之间;样本涵盖了全国东、中、西部的省份;44.5%的调研对象使用公共图书馆的频率在平均每月 1—9 次之间。

2.3 LLM 的选择

本研究选择了目前国内外相对典型的 LLM,来系统评测 LLM 对公共图书馆的感知水平,包括 OpenAI 公司开发的 GPT-4o-mini、深度求索公司开发的 DeepSeek-V2 和百度公司开发的文心一言。选择原因如下:(1)头部公司的代表性,OpenAI 和百度作为国际、国内领先的科技机构,在参数量、性能、用户量方面具有鲜明的优势;(2)推理成本较低,DeepSeek-V2 不仅在性能上直逼最先进的大语言模型 GPT-4,更是大幅度降低了推理成本,被开发者广泛应用于各类产品^[50];

(3)调用便捷性,三个 LLM 都支持 API(Application Programming Interface)调用,可以集中批量获取生成数据。基于以上考虑,本研究选取这三个 LLM 作为研究对象。

2.4 评估标准

本研究通过计算标准差来测量不同 LLM 在公共图书馆价值感知上的差异。采取平均绝对误差(MAE_{LLM-馆员/用户})和平均误差(ME_{LLM-馆员/用户})两个关键指标,来评估 LLM 与用户或馆员在公共图书馆感知水平上的差异。

平均绝对误差计算的是 LLM 评分与馆员或者用户评分之间的误差绝对值的平均值,这一指标能够反映感知差异的总体规模,而不受误差方向的影响,旨在量化 LLM 与馆员或者用户之间图书馆价值感知差异的大小。计算公式为:

$$MAE_{LLM-馆员/用户} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}| \quad \text{公式(1)}$$

平均误差则直接计算 LLM 评分和馆员或用户评分之间的误差平均值。这一指标尽管会受到正负误差相互抵消的影响,但是能够反映 LLM 的评分倾向,即相较于馆员或者用户群体在图书馆价值感知水平上是偏高还是偏低。计算公式为:

$$ME_{LLM-馆员/用户} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}) \quad \text{公式(2)}$$

由于本研究共测评了三个 LLM,所以 n 的值为 3。其中, y_i 代表第 i 个($i \leq 3$) LLM 对于某一服务类型问题项评分的平均值, \hat{y} 代表馆员或者用户群体对于某一服务类型问题项评分的平均值。

3 研究结果

3.1 馆员和用户对公共图书馆的价值感知

从六种图书馆价值的平均分值得来看,馆员对公共图书馆价值的感知水平平均高于用户和 LLM(见图 1),对各类价值感知的平均分都在 4.00 分左右,说明馆员对公共图书馆在各类服务职能上的表现都较为认可。其中,馆员对“文化服务”“社会教育”“阅读服务”的感知平均分均超过了 4.00 分,“文化服务”的认可程度最高(4.11 分),

“健康服务”的感知平均分则最低(3.82分)。用户对于各类价值的感知平均分均介于3—4分之间,意味着用户对公共图书馆的价值表示基本认可,但仍有一些期望未被完全满足。其中,图书馆

的“社会教育”价值受用户的认可度最高(3.89分),其次为“文化服务”(3.75分)，“健康服务”同样最受用户的认可(3.41分)。

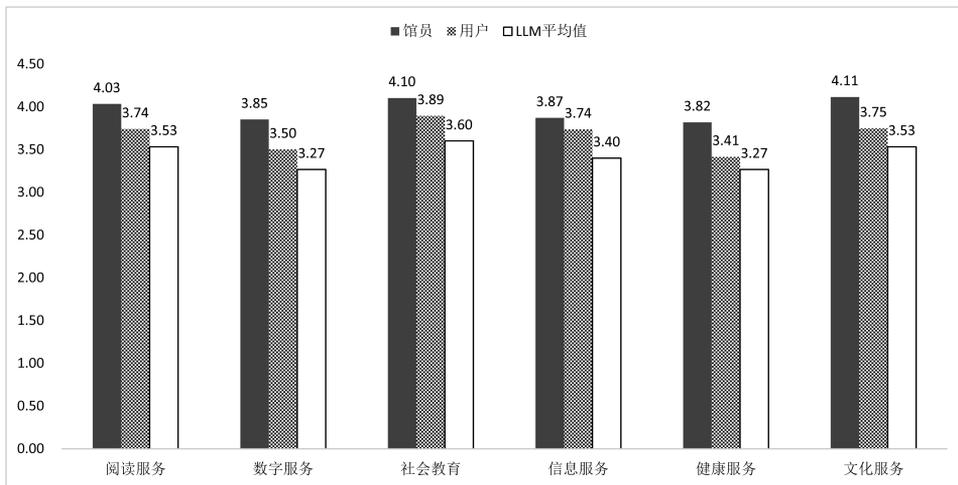


图1 馆员-用户-LLM三者对公共图书馆的价值感知平均分

在各类图书馆价值的具体题项中,馆员在“社会教育”下的分值极差最大,对于“弘扬志愿服务精神”这一题项给了最高的分值(4.65分),认为这一服务基本做到,且有接近于完美的服务效果;馆员针对“规范和教育公民的网络用语”这一题项则给了最低的分值(3.41分),反映了馆员对不同社会教育服务的认可度存在显著差异。用户评价最高的是公共图书馆作为“毕业/辞职之后的自习空间”这一“社会教育”价值(4.57分),评价最低的则是“健康服务”类型下的“向民众科普新冠等健康问题”这一题项(3.16分)。与馆员评价中“文化服务均分最高但具体题项最高分属于社会教育类”的差异不同,用户对公共图书馆价值的整体评价与具体题项评价之间具有较高的一致性。

3.2 LLM对公共图书馆的价值感知

研究表明,三个LLM对公共图书馆价值均持基本认可的态度,所有问题回答的平均分值介于3—4分之间。具体而言,文心一言对公共图书馆的平均评分最高(3.83分),其次是DeepSeek-V2(3.40

分),而GPT-4o-mini的平均评分最低(3.07分)。

进一步分析三个LLM的标准差,如图2所示,三个LLM对于公共图书馆“数字服务”的感知差异最大,标准差为0.7分,其中文心一言和DeepSeek-V2对于“数字服务”的感知存在显著分歧。其次是“健康服务”,标准差为0.64分,文心一言和GPT-4o-mini都对这一服务类型给予了最低评分,而DeepSeek-V2则给了最高评分。相比之下,“社会教育”的标准差最小(0.4分),反映出三个LLM对于公共图书馆“社会教育”价值上的感知趋于一致,且“社会教育”的平均得分(3.60分)是六类服务中最高的。

具体到各类公共图书馆价值的评分方面,文心一言对“健康服务”的评分最低(3.00分),但对其他五种价值的评分都达到了4.00分,表明其认为公共图书馆能够在这五类服务中体现较好价值。

与文心一言相反,DeepSeek-V2对“健康服务”给出了最高的平均分(4.00分),且五道题目均得到同样高分;评分最低的则是“数字服务”

(2.60分),尤其是对于“帮助老年人解决手机使用难题”“给小朋友提供科技互动和创客培训”“在乡村地区推广信息通讯技术”这三个具体题

项,DeepSeek-V2认为公共图书馆基本没有做好相关服务(2.00分)。

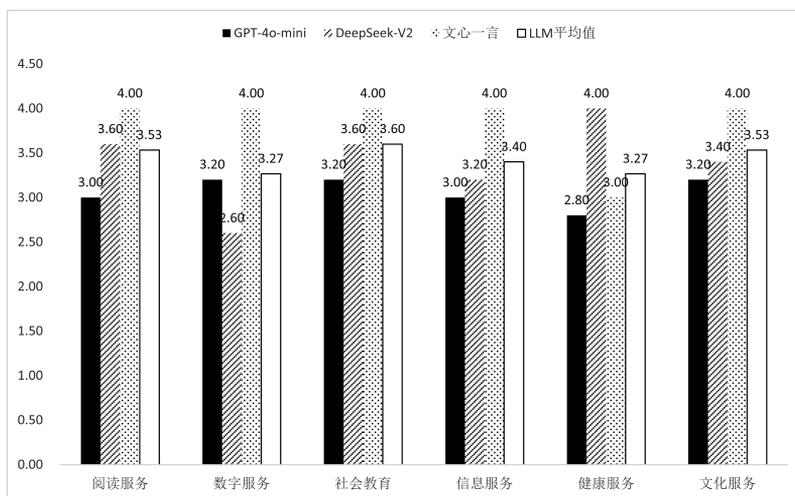


图2 不同 LLM 对公共图书馆的价值感知平均分

GPT-4o-mini 对我国公共图书馆价值的总体感知评分最低,如图3所示,其中有26道题(83.8%)的评价都是3分,包括了“信息服务”和“阅读服务”的10道题项,认为公共图书馆所发挥的效果可有可无。该模型对于“数字服务”“文化服务”“社会教育”三类的评分稍高,但也仅为3.20分。与文心一言一致,GPT-4o-mini 同样对“健康服务”的评分最低(2.80分),特别是对于“因地制宜做好地域性疾病的宣传”这一服务,GPT-4o-mini 认为公共图书馆基本没有做到。

综上所述,三个 LLM 对公共图书馆价值的感知存在一致性,也存在差异性。共性在于所有问题的回答平均分均在3—4分之间,反映出这三个 LLM 对公共图书馆中等偏上的认可程度。但差异性也很明显,从图3可得,国内大模型文心一言和 DeepSeek-V2 的评分多数集中在4.00分,针对大多数题目它们认为公共图书馆基本能够发挥作用。国外大模型 GPT-4o-mini 对我国公共图书馆价值的感知则相对较低,认为公共图书馆的服务效果在很大程度上是可有可无的。

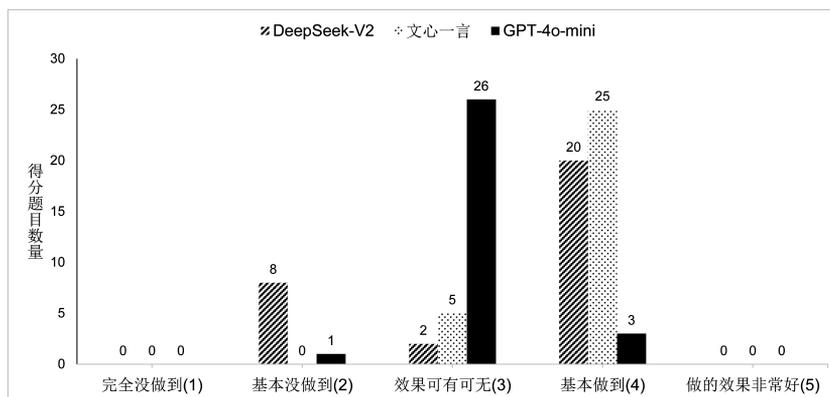


图3 LLM 对30道题目的评分分布

3.3 LLM 与馆员和用户的感知差异

在 LLM 与馆员和用户的感知比较方面,所有公共图书馆服务价值的 ME 得分均为负值(见图 4),表明相较于馆员和用户群体,LLM 对公共图书馆价值的感知评分是偏低的。此外,所有价值类型下,馆员的 MAE 值都大于用户的 MAE 值,说明 LLM 与馆员的公共图书馆价值感知差异大于与用户的差异,可见 LLM 对公共图书馆的价值感知更接近于用户而非馆员。

进一步分析不同价值类型的具体评分,如

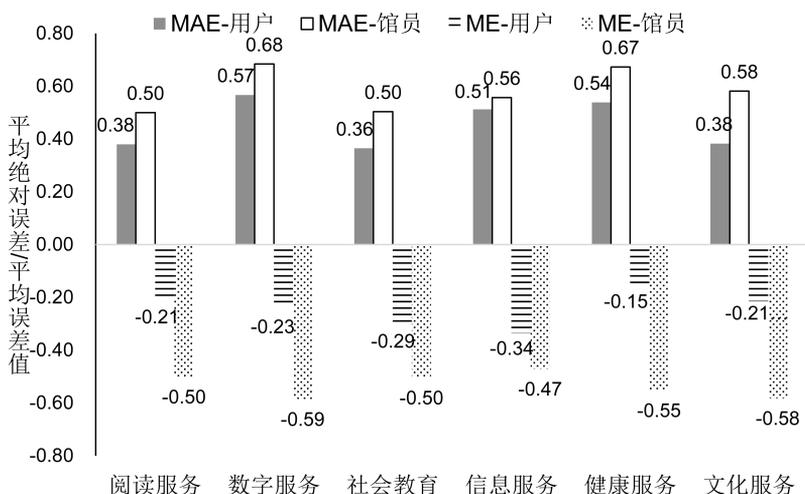


图4 LLM 与用户/馆员的公共图书馆价值感知差异值

如果将三个 LLM 视为整体,以三者评分均值代表 LLM 对公共图书馆的感知水平,一定程度上可以揭示 LLM 对公共图书馆价值的共感知规律。然而,这种分析可能忽略了不同 LLM 的特性,为此本研究进一步分析了单个 LLM 与馆员和用户之间的感知差异。

从整体评分均值的比较中可得,无论是与馆员还是与用户进行比较,文心一言在三个 LLM 中的 MAE 值都是最小的,最为接近人类主体对公共图书馆的价值感知。相反,GPT-4o-mini 与馆员、用户的 MAE 值都是最大的,显示出它与人类主体的公共图书馆价值感知差异最为显著。在具体差异大小方面,文心一言与馆员的 MAE 值小于与用户的 MAE 值,表明文心一言对公共图书馆的价

值感知更接近于馆员;其他两个 LLM 的感知水平则与用户更为接近。从 ME 值来看,只有文心一言与用户的比较结果为正数,表明只有文心一言对于公共图书馆价值的感知水平总体上高于用户,而其他两个 LLM 的公共图书馆感知水平则普遍低于用户和馆员。

在不同价值类型的深入分析中,如表 2 所示,三个 LLM 与馆员、用户的 ME 值大多数为负值,但文心一言和 DeepSeek-V2 在特定价值类型上显示出正值。其中,与用户的比较中,文心一言在对“健康服务”外的其他五类服务的价值感知上,ME 值均为正数,表明对于这五类服务,文心一言比用户更为认可公共图书馆的价值。在与馆员的比较中,文心一言只有在“数字服务”和“信息服

务”的 ME 值为正数,意味着对于这两类服务文心一言比馆员更能感知到公共图书馆的价值。DeepSeek-V2 只有在“健康服务”类型中的 ME 值为正数,对公共图书馆的“健康服务”价值感知高

于馆员和用户。GPT-4o-mini 在所有价值类型上的感知评价均低于用户和馆员,其中在“阅读服务”上与馆员、用户的感知差异最大,而在“数字服务”方面的感知差异最小。

表2 单个 LLM 与馆员、用户的 ME 值

价值类型 对比主体	阅读服务	数字服务	社会教育	信息服务	健康服务	文化服务	平均值
文心一言-馆员	-0.03	0.15	-0.10	0.13	-0.82	-0.11	-0.13
文心一言-用户	0.26	0.50	0.11	0.26	-0.41	0.25	0.16
DS-馆员	-0.43	-1.25	-0.50	-0.67	0.18	-0.71	-0.57
DS-用户	-0.14	-0.90	-0.29	-0.54	0.59	-0.35	-0.27
GPT-馆员	-1.03	-0.65	-0.90	-0.87	-1.02	-0.91	-0.90
GPT-用户	-0.74	-0.30	-0.69	-0.74	-0.61	-0.55	-0.61

4 讨论

本研究围绕公共图书馆的六种基本服务价值,设计了 30 道测试题目,首先探究了 LLM 对于公共图书馆价值的感知,其次以馆员和用户的感知作为基准,分析了 LLM 对公共图书馆价值的感知差异。

4.1 从 LLM 感知看公共图书馆的核心价值

本研究发现,公共图书馆的“社会教育”“文化服务”“阅读服务”价值得到了 LLM、馆员和用户三者的一致认可,而“健康服务”价值的认可程度在三者中均为最低。关于我国公共图书馆的媒介形象,研究表明公共图书馆呈现出“心灵避风港的知识天堂”“边缘化的阅读场所”“融合共享的文化服务平台”“工作生活相融合的第三空间”^[51],以及“公共文化服务体系搭建者”“公共文化承载者”“全民阅读保障者”^[35]等形象。这些形象揭示了“社会教育”“文化服务”“阅读服务”在公共图书馆服务中的核心地位,显示出 LLM 对公共图书馆传统核心价值的感知较为准确。

尽管公共图书馆开展“健康服务”的必要性和价值已经得到了充分的论证^[17],但是实际调查研究指出,我国公共图书馆在该领域的服务起步晚,尚处于实践探索阶段,虽然取得了一定的成效,但

尚未形成完整、成熟的服务体系^[52],且主要局限于我国各省级公共图书馆^[53]。由此可见,“健康服务”在我国公共图书馆的整体服务架构中尚未占据核心地位。这与本研究的发现也相呼应,LLM、馆员和用户一致对公共图书馆的健康服务评分最低,反映了公共图书馆在健康服务领域的实际表现与公众期望之间的差距。

具体到单个 LLM 的感知,文心一言和 GPT-4o-mini 都对公共图书馆的“数字服务”赋予了最高的评分,而该服务类型在馆员和用户的评分中均位于倒数第二位。这种差异可能源自于 LLM 对于新兴技术趋势的敏感性,以及它们在处理和分析数据时对数字服务的侧重。馆员和用户可能更加关注图书馆的传统服务价值,而文心一言和 GPT-4o-mini 的感知则提示了公共图书馆的服务价值应该是多元化和动态发展的,以不断适应社会的需求和价值观。

4.2 LLM 对公共图书馆的价值感知水平

4.2.1 不同 LLM 之间的感知差异

本研究所测试的三个 LLM 对公共图书馆服务价值的感知整体上呈现中等偏上的认可程度,但存在显著差异。国内大模型对多类服务价值持较高评价,而国外大模型则表现出相对较低的评价。这些差异可能与 LLM 的开发背景、文化差异以及

对图书馆服务价值的不同理解有关。本研究的30道测试题项的来源之一是我国公共图书馆的优秀实践案例,且以中文语境呈现,国内大模型在语义理解和价值判断上具有本土化优势;相反,国外大模型使用跨语言检索,所参考的资料中包括英文语料,因此在评估时的表现与国内大模型有所差异。

在六种服务价值类型上,“数字服务”的感知差异最为显著,“社会教育”则在所有模型中得到了较为一致的高评价。为了进一步探究三个LLM评分差异的原因,本研究以“数字服务”类型下评分差异最大的题项为例,对每个LLM的评分理由进行分析,由于LLM回答字数较长,表3只呈现理由的要点。

表3呈现了三个LLM对于该题项基本做到(4分)、效果可有可无(3分)、基本没做到(2分)这三种评分的理由。三个LLM都对公共图书馆在老年人智能手机培训方面进行了基本的肯定,尽管GPT-4o-mini对图书馆该类服务价值评价最低,但也承认了图书馆“发挥了一定的作用”,这种一致性反映了公共图书馆在“数字服务”类的基础价值。此外,三个LLM也审视了服务中存在的问题

和挑战,文心一言强调了全国范围内公共图书馆在服务趋势、多样化服务、资源投入和社区合作等方面的积极进展,并关注到了服务的地区差异问题;DeepSeek-V2的评分依据则更侧重于服务的普及性、持续性和系统性不足,强调服务体系构建的重要性;GPT-4o-mini则从老年人个体和社会层面分析了服务效果的局限性。

由此可见,LLM对公共图书馆的服务价值有着比较全面的理解,不仅能认识到不同地区和层级图书馆存在的服务水平差异,而且能够较为公正地评价公共图书馆在服务提供上的优势和不足。

4.2.2 LLM与人类主体之间的感知差异

研究发现,馆员对公共图书馆价值的认可度最高,而LLM对公共图书馆价值的感知评分最低。可能的解释是馆员作为图书馆服务的供给方,在评价公共图书馆价值时,更多从绩效角度出发,即关注图书馆的投入、产出及效率;用户虽然直接体验服务,但是评价更多地受到个人体验的影响;而LLM在一定程度上模拟了这种个性化的感知过程,因为它通过训练数据学习到的信息更接近普通用户的感知模式。

表3 “数字服务”评分差异最大题项的回答

数字服务:老年人不会使用智能手机,在数字社会中遇到了很多困难。帮助老年人解决手机使用难题,教他们如何使用智能手机,您认为图书馆发挥作用了吗?
<p>文心一言:基本做到</p> <p>1. 全国性趋势:越来越多的公共图书馆开始关注到老年人在使用智能手机上的困难,并纷纷采取措施来帮助老年人跨越“数字鸿沟”。2. 多样化服务:许多公共图书馆不仅提供基础的智能手机使用教程,还开设了专门的培训班、讲座或工作坊。3. 资源投入:不少公共图书馆加大了在设备、人力和资金等方面的投入。4. 社区合作:许多公共图书馆积极与社区、老年协会等组织合作,共同开展老年人智能手机培训活动。5. 地区差异:由于地区经济发展水平和文化背景的差异,不同地区的公共图书馆在服务质量和效果上仍存在一定的差距。然而,从整体上看,大部分公共图书馆都能够提供相应的服务或功能,并取得了基本的效果预期,但仍有进一步提升的空间和潜力。</p>
<p>DeepSeek-V2:基本没做到</p> <p>图书馆有责任和能力为老年人提供必要的帮助和支持,以缓解他们在数字生活中的不便。当前图书馆做的情况:1. 提供智能手机使用培训课程:部分大城市图书馆已经开始提供此类课程,但覆盖面和频率有限。2. 设立咨询服务台:一些图书馆设有咨询服务台,但专门针对智能手机使用的服务并不普遍。3. 制作和提供教学资料:少数图书馆提供了相关资料,但资源有限且更新不及时。4. 开展一对一辅导:个别图书馆尝试了这种服务模式,但尚未形成规模和系统。</p> <p>综上所述,虽然图书馆在帮助老年人学习智能手机使用方面已经有所行动,但在服务的普及性、持续性和系统性方面仍有较大的提升空间。因此,评价选项“基本没做到”是合理的,因为图书馆在帮助老年人解决手机使用难题方面尚未充分发挥其应有的作用。</p>

GPT-4o-mini:效果可有可无

1. 个体差异:老年人的学习能力和适应能力存在差异,图书馆的帮助效果可能因人而异。2. 资源限制:如果图书馆的资源不足,如志愿者数量、教学材料、设备等,帮助的效果可能大打折扣。3. 持续性问题:如果老年人没有持续的支持和训练,学习效果可能会减弱。4. 技术更新迅速:智能手机技术和应用程序更新较快,导致之前的学习效果变得不再适用。5. 社会支持因素:如果周围没有人能够给予帮助和鼓励,图书馆的帮助可能也会显得有限。综上所述,虽然图书馆在帮助老年人使用智能手机方面可能发挥了一定的作用,但由于上述种种原因,其效果并不是绝对有效,可能在特定情况下表现为“可有可无”。

为了进一步评价 LLM 对公共图书馆的感知水平,本研究选择公共图书馆馆员和用户的感知作为基准,并以平均绝对误差 (MAE) 和平均误差 (ME) 两个指标量化感知水平的差距。研究发现,不同 LLM 表现出不同程度的公共图书馆价值感知水平。具体来说,文心一言与馆员和用户的感知最为接近,一定程度上可以说明文心一言能更为深入地理解公共图书馆服务的价值,尽管在特定服务类型上可能存在过于乐观的感知,例如除了“健康服务”外的其他五类服务,文心一言的认可度都要高于用户。GPT-4o-mini 则表现出了与人类主体最大的感知差异。

4.3 LLM 与人类感知差异对公共图书馆的影响

本研究通过实证测试揭示了 LLM 与馆员和用户之间存在一定的公共图书馆价值感知差异,且在所有考察的服务类别中,LLM 对公共图书馆服务价值的评分均值都低于馆员和用户,表明 LLM 相较于人类对公共图书馆价值的认可度较低。在 LLM 日益成为协同决策智能体的背景下,这一感知差异可能导致用户在面对六类服务需求时,产生减少选择和利用公共图书馆的倾向。

值得一提的是,研究发现 LLM 与用户在“信息服务”的 ME 值最高,而在“健康服务”的 ME 值最小。这表明,LLM 协同人类决策的影响在“信息服务”领域中最显著,而在“健康服务”领域中最轻微。长期以来,公共图书馆作为“信息中心”^[45]的使命都是图书馆信息职业的传统感知,图书馆情报学的基本问题就是“如何保障信

息有效查询和有效获取”^[47],无论是业界实践还是学界研究,都将“信息服务”视为公共图书馆最为核心的社会职能。当 LLM 对公共图书馆“信息服务”的价值感知评分低于用户时,这可能会挑战公共图书馆作为信息中心的传统角色。在 LLM 的辅助决策之下,用户可能会更倾向于使用其他信息渠道,如在线数据库、搜索引擎或社交媒体,而不是依赖公共图书馆。

4.4 AIGC 背景下图书馆信息职业的应对策略

面对 LLM 与人类感知差异所带来的影响,图书馆信息职业必须积极采取措施,以适应 LLM 参与决策的新环境,促使公共图书馆服务的价值得到更为广泛的认可和利用。基于研究发现,本文提出以下建议:

(1) 积极参与 LLM 生态建设:图书馆信息职业应进一步融入 LLM 的生态体系,不仅仅是作为新技术的采纳者,更应成为 LLM 发展的促进者。通过发布公开数据集、构建评测任务等方式,为 LLM 提供丰富的业务场景和数据支持,从而帮助 LLM 更准确地理解和评估公共图书馆服务的价值。

(2) 赋能 LLM 服务:应充分发挥图书馆知识库与语料库资源优势^[54],将自身打造成用户获取和利用 LLM 服务的高地。一方面,以图书馆业务需求为导向,将 LLM 整合到图书馆的业务流程中,使 LLM 为馆员的办公、业务和学术研究提供智慧化支持;另一方面,以 LLM 为核心扩展图书馆的服务应用,为读者提供统一的接口以及更为个性化和高效的服务^[55]。

(3) 建立图书馆服务评价的新标准;应考虑建立更加全面、客观的服务评价体系,不仅包括用户满意度、服务效率等传统绩效和成效指标,也应考虑 LLM 的感知结果,在此基础上建立的评价体系能够更好地助力图书馆的未来事业发展。

5 结论

本研究在充分调研公共图书馆的相关文献研究和实践工作案例的基础上,提炼出“阅读服务”“信息服务”“数字服务”“社会教育”“文化服务”“健康服务”这六类公共图书馆服务价值,每类服务价值下均设计了5道问题,作为公共图书馆价值感知测试题;分别对 GPT-4o-mini、文心一言和 DeepSeek-V2 三个 LLM 进行了测评,同时以公共图书馆馆员和用户的问卷调查结果作为基准,比较分析 LLM 对我国公共图书馆价值的感知水平。研究发现:(1) LLM、馆员和用户都对“社会教育”“文化服务”“阅读服务”持高度认可,“健康服务”则被三者赋予最低的评价,这与公共图书馆的传统核心价值相契合。(2) 不同 LLM 之间存在感知差异,国内大模型对多数服务价值类型持较高评价,而国外大模型则表现出相对较低的评价,这一发现强调了大模型训练数据的地域性和文化相关性对于评估特定领域服务价值的重要影响。(3) LLM 与馆员和用户之间存在公共图书馆感知差异,LLM 对公共图书馆所有服务价值的认可度普遍低于人类。尤其在“信息服务”领域,LLM 的感知评分显著低于用户,这可能会挑战公共图书馆作为信息中心的传统角色。基于以上发现,本研究围绕 LLM 生态建设、服务赋能、评价标准三个方面,对图书馆信息职业应对 LLM 与人类的价值感知差异提出了建议。

从理论价值来看,本研究突破了图书馆价值研究惯常聚焦的“馆员-用户”二元认知框架,率先引入 LLM 作为新型认知主体,将技术人工物的价值感知纳入图书馆价值研究视域。同时,本研究构建了一套测量公共图书馆价值感知的数据集,以馆员和用户的图书馆感知作为基准,对 LLM 的

图书馆感知水平进行测评。这一努力有助于从更为全面的视角来洞察我国公共图书馆在 AIGC 背景下的角色和地位,以及 LLM 对公共图书馆传统价值定位、公众利用决策的可能影响,并启发后续研究探寻可能的影响机制、表现形式。从实践价值来看,本研究对于 AIGC 背景下图书馆信息职业的未来进路具有重要的参考作用。在智慧图书馆的蓬勃发展中,以 LLM 赋能图书馆智慧服务并将图书馆打造为用户接触 LLM 的重要平台,形成图书馆价值导向的 LLM 服务,有助于引导用户正确理解和使用 LLM,减少用户对 LLM 的盲目依赖,提高用户对图书馆价值的感知和认同。

当然,本研究也存在不足之处。首先,以馆员和用户的评测结果作为基准,可能会受到个体经验和偏好的影响;其次,研究只选取了三个 LLM 进行评测,样本数量的有限性可能会影响结果的代表性。此外,这些 LLM 生成的结果未被多次重复检验,可能受到模型性能稳定性的影响。未来研究可以考虑更多的 LLM,以及更为广泛、异质性的用户群体,以提高研究结果的代表性和普适性。

参考文献

- 1 程焕文. 从城市温度、新闻热度看公共图书馆平等权利的实现程度——对“读者留言东莞图书馆”新闻热点的思考[J]. 图书馆论坛, 2020(10):1-8.
- 2 陈忆金,曹树金. 用户中心视角下公共文化服务质量评价研究[J]. 图书情报工作, 2019(17):60-68.
- 3 Ladhari R, Morales M. Perceived service quality, perceived value and recommendation: a study among Canadian public library users[J]. Library Management, 2008, 29(4/5):352-366.
- 4 OpenAI 正式进军搜索,发布 SearchGPT:专挑谷歌不会的问题演示[EB/OL]. [2024-08-15]. https://www.thepaper.cn/newsDetail_forward_28204231.
- 5 Liu Y, et al. Trustworthy LLMs: a survey and guideline for evaluating large language models' alignment[J]. arXiv preprint arXiv:2308.05374, 2023.

- 6 Girotra K, et al. Ideas are dimes a dozen: large language models for idea generation in innovation [J]. Available at SSRN 4526071, 2023.
- 7 Liu S, et al. Large language models as evolutionary optimizers[C]//2024 IEEE Congress on Evolutionary Computation (CEC). IEEE, 2024:1-8.
- 8 李超平,叶斌.基于成本效益分析的公共图书馆经济价值研究[J].中国图书馆学报,2012(2):12-20.
- 9 周珊.以“投资回报率”评估图书馆的价值——国际图联《用数据回顾过去十年图书馆的投资回报率》报告解读与启示[J].大学图书馆学报,2021(2):10-19.
- 10 Aabø S. Are public libraries worth their price? A contingent valuation study of Norwegian public libraries[J]. New Library World, 2005, 106(11/12):487-495.
- 11 Stejskal J, et al. The perceived value of book borrowing services is stationary in the time of Covid-19: empirical evidence from the Municipal Library in Prague[J]. Journal of Librarianship and Information Science, 2023, 55(4):948-959.
- 12 Gómez-Zapata J D, et al. Valuing cultural public goods in times of pandemic: what happened to the libraries? [J]. Journal of Librarianship and Information Science, 2023;09610006231196345.
- 13 Noh Y. A study of the evaluation of a library's economic value based on the users' and librarians' perceptions [J]. The Electronic Library, 2020, 38(3):561-596.
- 14 Stenstrom C, et al. A review exploring the facets of the value of public libraries [J]. Library Management, 2019, 40(6/7):354-367.
- 15 周潇依,金武刚.中外公共图书馆“场所”价值的公众感知比较研究——以“IFLA 绿色图书馆奖”入围公共图书馆为样本[J].图书与情报,2022(3):31-44.
- 16 Noh Y, Lee S Y. An evaluation of the library's educational value based on the perception of public library users and librarians in Korea [J]. The Electronic Library, 2020, 38(4):677-694.
- 17 周晓英,于小宁.公共图书馆健康服务的价值感知和实践探索[J].中国图书馆学报,2022(4):90-103.
- 18 Noh Y, et al. A study on comparing the perception between librarians and users about libraries' value [J]. International Journal of Knowledge Content Development & Technology, 2018, 8(2):67-100.
- 19 Noh Y. A study on the evaluation analysis of the library's social values[J]. Journal of Librarianship and Information Science, 2021, 53(1):29-49.
- 20 吴莉,等.用户视角下公共图书馆影响力评价研究——以广州图书馆为例[J].国家图书馆学刊,2023(3):24-35.
- 21 Creaser C. Assessing the impact of libraries—the role of ISO16439 [J]. Information and Learning Science, 2018, 119(1-2):87-93.
- 22 银晶,冯玲.国际标准 ISO16439:2014《信息与文献——图书馆影响力评估的方法与流程》解读与分析[J].山东图书馆学刊,2017(2):66-70.
- 23 Vakkari P, Serola S. Perceived outcomes of public libraries [J]. Library & Information Science Research, 2012, 34(1):37-44.
- 24 Noh Y. A study on the perception of librarians of the impact that libraries have on society [J]. International Journal of Knowledge Content Development & Technology, 2016, 6(2):109-147.
- 25 唐琼,蓝丽婷.我国主流媒体报道语境下的公共图书馆影响力表现维度研究[J].图书情报知识,2021(3):74-87,13.
- 26 肖鹏,廖书语.官方媒体视域下的档案馆功能及其演变:以《人民日报》70年的新闻报道为中心[J].档案学研究,2020(1):111-118.
- 27 肖鹏,赵庆香.公共图书馆在中国减贫事业中的作为、贡献与经验(1978—2020)[J].中国图书馆学报,2023(1):38-57.
- 28 周丹燕,唐琼.变化理论视角下公共图书馆服务项目成效评估研究[J].图书情报工作,2022(16):48-56.
- 29 周九常.公共图书馆社会形象体系模型构建[J].图书馆理论与实践,2013(5):11-15.
- 30 张麒麟.图书馆形象的历史嬗变及其在元宇宙

- 宙中的构建[J]. 国家图书馆学刊, 2022(4): 50-57.
- 31 冯昌扬,等. 听听人工智能的声音: 社会符号学视角下人工智能对图书馆员职业的感知[J]. 图书馆论坛, 2025(2): 72-81.
- 32 周樱格,等. 图书馆吉祥物体验式营销的实践与探索——以金陵图书馆为例[J]. 图书情报工作, 2021(14): 74-81.
- 33 文捷敏,等. 文旅融合背景下新型公共文化空间的形象感知[J]. 图书馆论坛, 2025(1): 29-39.
- 34 李勇,等. 文旅融合下的省级公共图书馆公众形象感知与质量提升[J]. 图书馆论坛, 2021(8): 108-119.
- 35 张守卫, 杨小丽. 2000年以来我国省级公共图书馆媒介形象及其特征分析——以报纸报道为数据源的文本分析[J]. 图书情报工作, 2024(7): 57-68.
- 36 Duxfield A, Liew C L. Libraries in contemporary science fiction novels: uncertain futures or embedded in the fabric of society? [J]. Journal of Documentation, 2023, 79(3): 546-566.
- 37 何帆, 曾一昕. 网络新闻中的图书馆社会形象[J]. 图书馆论坛, 2021(5): 10-19.
- 38 Frank M C. Baby steps in evaluating the capacities of large language models [J]. Nature Reviews Psychology, 2023, 2(8): 451-452.
- 39 Yiu E, et al. Transmission versus truth, imitation versus innovation: what children can do that large language and language-and-vision models cannot (yet) [J]. Perspectives on Psychological Science, 2024, 19(5): 874-883.
- 40 Han S J, et al. Inductive reasoning in humans and large language models [J]. Cognitive Systems Research, 2024, 83: 101155.
- 41 Ziems C, et al. Can large language models transform computational social science? [J]. Computational Linguistics, 2024, 50(1): 237-291.
- 42 Wan Y, et al. "Kelly is a warm person, Joseph is a role model": gender biases in LLM-Generated reference letters [J]. arXiv preprint arXiv: 2310.09219, 2023.
- 43 Manvi R, et al. Large language models are geographically biased [J]. arXiv preprint arXiv: 2402.02680, 2024.
- 44 Motoki F, et al. More human than human: measuring ChatGPT political bias [J]. Public Choice, 2024, 198(1): 3-23.
- 45 The IFLA-UNESCO Public Library Manifesto 2022 [EB/OL]. [2024-05-24]. <https://www.ifla.org/public-library-manifesto/>.
- 46 图书馆服务宣言 [EB/OL]. [2024-09-16]. https://www.lsc.org.cn/cns/html_pass/contents/1676363541657/1703426584604184576.html.
- 47 于良芝. 图书馆情报学概论 [M]. 北京: 国家图书馆出版社, 2016: 3, 251-252.
- 48 周晓英. 健康服务: 开启公共图书馆服务的新领域 [J]. 中国图书馆学报, 2019(4): 61-71.
- 49 柯平,等. 面向社会化挑战的图书馆专业化学理依据与实践探索 [J]. 中国图书馆学报, 2023(2): 15-28.
- 50 Liu A, et al. Deepseek-v2: A strong, economical, and efficient mixture-of-experts language model [J]. arXiv preprint arXiv: 2405.04434, 2024.
- 51 张蔚林, 陈幼华. 当代图书馆的媒介形象建构——以《人民日报》2009—2018年报道为例 [J]. 图书馆论坛, 2020(1): 96-104.
- 52 姚丽琴,等. 我国公共图书馆健康信息服务研究 [J]. 图书馆工作与研究, 2024(8): 83-91.
- 53 袁梦群, 严贝妮. 我国省级公共图书馆健康信息服务调查与启示 [J]. 图书馆理论与实践, 2019(2): 10-14.
- 54 吴建中,等. 面向未来的图书馆与社会 [J]. 中国图书馆学报, 2021(2): 4-28.
- 55 上海图书馆(上海科学技术情报研究所),等. 智慧图书馆大模型创新与应用白皮书(征求意见稿) [EB/OL]. [2024-08-25]. <https://www.calsp.cn/2023/09/26/bulletin-202309-03/>.
- (陈慧彤 中国人民大学信息资源管理学院情报学专业 2023 级博士研究生, 郑尔特 谢菲尔德大学信息学院情报学专业 2024 级博士研究生, 闫慧 教授 中国人民大学信息资源管理学院, 赵雪锋 馆员 首都图书馆)

收稿日期: 2024-11-30