

·图书馆与图书馆事业·

“图书馆+DeepSeek”的应用路径研究*

吴若航 茆意宏

(1.南京农业大学信息管理学院 江苏南京 210031)

摘要:DeepSeek为图书馆智能化发展带来新契机,探究图书馆对于DeepSeek的应用路径,对推动图书馆AI应用的发展具有重要意义。文章在探析DeepSeek突破性特征的基础上,对DeepSeek的应用态势及应用挑战进行剖析,并归纳出“图书馆+DeepSeek”的应用路径。研究发现,“图书馆+DeepSeek”应用路径主要以“战略层-建设层-运营管理层-评价层”为核心,图书馆需做到明确认知与确定规划,完成技术系统、服务内容及馆员队伍的建设,做好人财物、技术系统、数据方面的运营管理工作,并从建设和效用两方面进行实时评价。

关键词:DeepSeek;图书馆;AIGC;生成式人工智能

中图分类号:G250.7 **文献标识码:**A **DOI:**10.11968/tsyqb.1003-6938.2025021

Research on the Application Pathway of "Library + DeepSeek"

Abstract DeepSeek presents new opportunities for the intelligent development of libraries. Exploring the application pathways of DeepSeek in libraries is of great significance for advancing AI applications in the field. Based on an analysis of DeepSeek's breakthrough features, this study examines its application landscape and challenges, ultimately summarizing the "Library + DeepSeek" application pathway. The findings reveal that this pathway is structured around four core layers: strategic planning, system construction, operational management, and evaluation. Libraries must establish a clear understanding of the technology and formulate precise implementation plans, complete the development of technical systems, service content, and staff teams, ensure effective management of resources, technical infrastructure, and data, and conduct real-time evaluations from both construction and utility perspectives.

Key words DeepSeek; library; AIGC; Generative Artificial Intelligence

2025年1月20日,我国的深度求索公司正式发布大模型——DeepSeek-R1,该款AI产品在算法效率、能耗成本及开源策略上的突破使其瞬间成为现象级产品,引发国际媒体和各行业的广泛关注。DeepSeek-R1发布一周后迎来用户爆发式增长,并迅速登顶多个国家的应用商店下载榜,其用户突破1亿仅用了7天,增速远超ChatGPT(2个月)、抖音(9个月)等知名互联网产品,创下了全球互联网产品用户增速的新纪录^[1]。DeepSeek拥有卓越的性能,在数学、代码和推理任务上的输出效果可与OpenAI-o1相媲美。自DeepSeek开源后,众多政府部门、企事业单位迅速宣布接入DeepSeek-R1,真正实现了人工智能(Artificial Intelligence, AI)应用的本地化部署,

极大程度上促进了我国AI应用的发展。

图书馆行业长久以来始终关注着信息技术的发展与应用,从互联网普及过程中出现的数字图书馆,到如今AI技术迅猛发展背景下的智慧图书馆,作为知识与文化的枢纽,图书馆正加速推进智能化转型。近几年,图书馆界已逐步引入各类AI技术优化服务效果,如通过云计算、生物识别、模式识别、机器学习、自然语言处理、神经网络等技术实现图书馆的智能化场馆建设、智能化服务、智能化资源建设以及智能化管理^[2]。部分图书馆也尝试借力生成式AI推动图书馆的智能化应用,如国家图书馆基于百度的“文心大模型”发布的智能问答系统“顾朝夕”,可为读者提供实时的AI生成内容服务^[3]。然而,以往图

*本文系国家社会科学基金重点项目“数智时代阅读服务转型研究”(项目编号:21ATQ003)研究成果之一。

收稿日期:2025-03-27;通讯作者:茆意宏(maoyh@njau.edu.cn);责任编辑:柴若熔

图书馆在智能化转型过程中受限于经费投入不足、技术壁垒、馆员 AI 素养有待提升等问题, 图书馆整体在 AI 领域的拓展性以及相关的前沿技术应用不足, 在个性化服务与深度交互方面仍存在瓶颈。在此背景下, 以降本增效、开源普惠为突破点的 DeepSeek 能够为图书馆行业的 AI 应用推进工作带来新的契机。基于此, 本文拟从 DeepSeek 的突破性特征分析出发, 剖析 DeepSeek 的应用态势与应用挑战, 构建出“图书馆+DeepSeek”的应用路径, 为我国图书馆推进 AI 应用的工作开展提供参考。

1 DeepSeek 的突破性特征

在 DeepSeek 问世之前, 国内的百度、阿里巴巴、腾讯等互联网公司也曾推出文心一言、通义千问、混元大模型等 AI 产品进行市场探索, 但均未能像 DeepSeek 那样成为大规模现象级产品。DeepSeek 之所以能够达到前人难以企及的高度, 主要原因在于其通过技术创新和市场策略实现了“更智能”“更亲民”“更普及”等多方面突破。

1.1 算法技术突破——降本增效

DeepSeek 系列作为大模型产品, 其关键技术核心仍绕不开算法、算力及数据三个层面。在算力层面, 虽然深度求索的母公司幻方早已在几年前就开始布局 AGI 的研发工作, 并早在 2019 年成立之初就开始集攒显卡, 成为早期国内能够达到 ChatGPT 算力要求量级(万张 A100 加速卡)的六家公司之一(其余五家分别是商汤科技、百度、腾讯、字节跳动和阿里巴巴)^[4], 但英伟达掌控全球大部分算力主权的形势仍无法撼动, 因此在算力层面难以实现真正突破。在数据层面, 以 DeepSeek-V3 为例^[5], V3 模型在预训练阶段使用了 14.8 万亿个高质量标记的 Token, 充分保障了训练数据的质与量, 但数据层面的工作更多是作为底层基础, 人类的数据资源是有限的, 数据层面的跨越级突破也较难实现。既然在算力和数据层面难以突破, DeepSeek 团队便通过优化与创新将模型的算法效率提升到极致, 最终成功在算法层面上取得突破性进展。

我们可以简单将效率理解为有效产出与投入成本的比值, 即“效率=产出/投入”, 因此提升算法效率

可以从两个方面改善, 一是提高产出, 二是压缩投入, DeepSeek 在这两个方面进行了无比细致的规划, 以达到降本增效的效果。在提高产出方面, DeepSeek 团队主要的算法创新及应用突破在于以下几点: (1) 优化后的混合专家模型(Mixture of Experts, MoE)架构, 该架构通过组合多个专家模型以处理复杂任务, 已被多款 AI 大模型使用, 但 DeepSeek 的创新之处在于细粒度专家划分、共享-路由专家策略及无辅助损失负载均衡策略, 让模型保持高效训练和动态平衡的同时, 显著提升性能。(2) DualPipe 流水线并行算法, 有效改善了流水线的停滞现象, 实现计算与通信阶段的重叠, 从而提升了 GPU 的利用率, 并降低了通信开销^[6]。(3) 多 token 预测(Multi-Token Prediction, MTP)技术, 尽管 MTP 技术是由美国的 Meta 公司提出的^[7], 但 DeepSeek 率先将 MTP 应用于模型训练中, 通过并行预测多个 token、增强预测能力并提升数据效率。在压缩投入方面, DeepSeek 团队引入了创新性的多头潜在注意力(Multi-head Latent Attention, MLA)机制, 该机制是由多头注意力机制(Multi-Head Attention, MHA)^[8]改进而来, 能够让模型通过更高效的注意力交互方式, 缓解了大模型在长序列、高维度上的计算瓶颈, 是 DeepSeek 能够实现低成本训练大模型的关键环节^[9]。除了 MLA 机制之外, DeepSeek 还创新了 8 位浮点数(Floating Point 8bit, FP8)混合精度训练框架, 以牺牲部分非重要精度为代价, 提升了计算速度, 有效提高了单位 GPU 小时的计算利用率, 极大降低了整体的训练成本, 对于需要更高精度的模块, DeepSeek 则保留了 BF16/FP32 精度, 平衡了效率与稳定性。

DeepSeek 的 V3、R1 模型训练成本在百万美元级别, 相较于动辄上亿美元训练成本的国外大模型, DeepSeek 通过算法优化取代算力堆砌成为核心竞争力。低成本是技术普及的基本要求, DeepSeek 的技术路线给更多的国家与企业展示了以算效为核心, 降低成本入局 AI 的可能性, 推动 AI 格局进入以追求高算效和高能效为主的新阶段^[10]。

1.2 市场策略突破——开源普惠

Meta 的首席 AI 科学家杨立昆对于 DeepSeek 的评价是“开源模型正在超越专有模型”, DeepSeek 能

够引发全球热潮的原因之一在于其选择了合适的市场策略。DeepSeek并不是首个开源大模型,部分国外企业也尝试过在开源市场进行探索(如2023年Meta发布的开源大模型LLaMA),但其实际市场效果远不如DeepSeek,究其原因在于以下几点。

首先,DeepSeek打破了人们先前的“开源=低配”的认知。以往的开源模型性能明显落后于闭源模型,而DeepSeek-R1开源模型的性能能够做到比肩甚至在某些模块上优于对标的闭源模型OpenAI-o1^[11],首次实现了开源模型性能对标闭源模型性能标杆,让大多数集体和个体在选择大模型产品时会更倾向于能进行本地化部署以保障数据安全的同时且性能不弱于闭源产品的DeepSeek。其次,DeepSeek在开源的基础上还选择了门槛极低的MIT协议,即允许用户自由使用、修改和对修改后模型的商业化,这意味着开发者的限制被极大降低,尤其是对于企业而言能够有效减少试错成本,从而推进应用快速落地。再次,DeepSeek在价格上具有显著优势,尤其是V3模型,其输入和输出价格远低于OpenAI的GPT-4o等高端模型^[12],再加上“夜间API调用大降价”策略^[13],可以让用户享受更经济、更流畅的使用体验,真正意义上实现了价格普惠,进一步提升了市场热情。此外,DeepSeek专注构建全栈式的开源生态,其在2024年2月下旬的开源周连续开源了FlashMLA、DeepEP、DeepGEMM、DualPipe、EPLP、3FS、Smallpond^[14],在公开训练代码、技术论文和研发细节的同时,还提供了模型蒸馏工具链,吸引了众多全球研发者参与优化,加快迭代速度。

开源生态能够形成巨大的虹吸效应,一旦其具备良好的规模,就能与闭源市场相抗衡^[15]。DeepSeek的开源策略也逐渐影响到了大模型的行业走向,越来越多的AI企业也开启了大模型的开源计划。如2025年2月,字节跳动、昆仑万维、百度文心一言等均推出开源模型^[16];2025年3月,阿里巴巴发布并开源了全新的大模型通义千问QwQ-32B,该模型的性能可与DeepSeek-R1相媲美,同时兼具成本低的特点^[17]。开源普惠的AI行业态势将进一步提升AI的普及率和渗透率,让AI大模型走进各行各业以及每个普通人的生活不再遥不可及。

2 DeepSeek的应用态势及应用挑战

2.1 DeepSeek在各行业的应用态势

DeepSeek-R1在发布后迅速“破圈”引发热议,高性能、低成本、本地部署安全性高等优势让DeepSeek在多行业呈现出跨领域渗透和生态重构态势,政府、金融、医疗、互联网、教育等行业陆续开始根据自身的需求与特色进行大模型的应用规划,并打造出了多元化、可落地的应用场景。目前各行业对于DeepSeek的应用态势聚焦于以下两个方面:

(1)利用DeepSeek提升业务效能。DeepSeek对于各行业的业务效能提升是最显著的应用态势,此类工作通常以推进DeepSeek通用场景落地为主,如对话式问答、智能检索、工作内容生成以及文本解读等,利用大模型提升原有业务质量与效率,落地难度较低^[18]。如广州“12345”政务热线系统接入DeepSeek之后,话务员可以在DeepSeek智能转化语音的协助下,快速梳理来电诉求,同时精准提取诉求热线的核心要点,对数据进行标注归类,实现“分得准、办得快”^[19];微信、百度、知乎、微博等互联网平台陆续宣布接入DeepSeek,通过AI搜索提升用户的使用体验^[20];众多央企通过DeepSeek的本地化部署创建了AI智能助手应用,实现了智能问答、智能办公等业务优化,有效提升了员工的工作效率^[21]。

(2)通过DeepSeek训练垂直领域应用。DeepSeek的本地化部署浪潮已经让多个行业结合本地数据集与知识库逐渐实现垂直领域深化,尤其在金融、医疗、工程设计等行业态势明显,此类垂直应用场景属于AI应用的“深水区”,落地难度较大,但潜在价值高。如中国邮政储蓄银行依托自有大模型“邮智”,本地部署并集成DeepSeek模型,构建出股权财报分析助手、AI财富管家、风险防控助手等应用场景^[22];南通市第一人民医院在接入DeepSeek模型后,首次将放射信息系统和病理系统与R1模型进行融合,构建“智能报告质控+病理符合判断”双保障体系,该功能使单病例复核效率提升400%^[23];长江设计集团基于DeepSeek的本地化部署,创新构建了防洪知识大模型、防洪决策大模型、“决策智脑”防汛指挥系统、“库区地质灾害”及“隧洞地质风险”等模块

构成的地灾数字孪生平台,将传统方案制定周期大幅压缩^[24]。

2.2 DeepSeek 在图书馆行业的应用态势

图书馆界时刻关注着信息技术的发展动态,在此次 DeepSeek 浪潮下,国内已有多家图书馆宣布接入了 DeepSeek,为用户提供前沿的智能化服务,其应用场景主要聚焦于智能问答、资源推荐、业务办理以及科研支持等。如扬州大学图书馆的 AI 馆员“图小智”全面搭载 DeepSeek-R1 满血版,通过对接图书馆专有知识库,覆盖图书馆全场景咨询服务,使问题精准直达,答复效率翻倍^[25];兰州大学图书馆的 AI 馆员借助 DeepSeek 算法,可从海量文献中精准匹配用户需求,让图书、期刊、论文一键直达,还可以根据用户的阅读习惯推荐匹配读物^[26];山东大学图书馆的 AI 馆员“山小图”依托 DeepSeek 系列模型,深度对接图书馆业务系统,实现图书续借、座位预约、入馆教育等具体业务的一站式服务集成,极大程度上方便了用户使用^[27];西北工业大学图书馆搭载了 DeepSeek 满血版的 AI 馆员能够完成一站式 AI 检索,并根据检索结果快速获取原文文献,还能够帮助用户进行文献精读,以及提供自动生成文献综述、开题报告等科研支持服务^[28]。除了上述 DeepSeek 在图书馆领域的通用应用场景之外,部分图书馆也开始通过 DeepSeek 训练垂直领域应用,如常州图书馆于 2 月 25 日发布的基于 DeepSeek 本地化部署的常州地方文化大语言模型,以“垂直深耕”为特色,提供地方文化资源的一站式知识检索和深度解读,相较于使用通用 AI 工具可以得到更加专业和精准的回答,这也标志着公共图书馆领域的 AI 融合实践迈入新阶段^[29]。

2.3 DeepSeek 的应用挑战

DeepSeek 的广泛应用案例反映了当下基于 AI 大模型应用的各行业智能化总体发展趋势。在部分成功案例的示范效应和各行业智能化转型的绩效压力的双重驱动下,不同行业在大模型部署工作中呈现出相似的“DeepSeek 狂热”态势,各类 AI 助手的部署浪潮迭起。但与此同时,由于部分机构单位缺乏大模型应用经验与系统性规划,导致其在部署 DeepSeek 之后并未有效提升相应的业务质量,反而可能造成资源浪费与安全问题,此类现象值得业界深入

思考。本文以图书馆行业为例,剖析图书馆在 DeepSeek 的应用过程中可能会面临的挑战,为后续的“图书馆+DeepSeek”应用路径构建提供应对参考。

2.3.1 形式主义与技术主义

图书馆在推进 DeepSeek 应用的过程中可能会出现形式主义与技术主义现象。部分图书馆在引入大模型应用时,可能会更关注智能化的表层效果,而非深度融入实际服务,导致技术应用流于表面,仅形成华丽的交互界面,未能真正提升用户使用率与使用体验。这种“形象工程”往往忽视了技术落地的核心价值,使得投入巨资打造的 AI 馆员沦为摆设,而无法做到实质性的服务效能提升。同时,当前互联网市场上的 DeepSeek 产品层出不穷,包括腾讯元宝、百度 AI 搜索等平台,均已接入完整版本的 DeepSeek,并向用户提供免费使用权限。在这一背景下,经费相对有限的图书馆若未能精准定位自身需求、合理评估实际条件,而仅是盲目跟风上线类似产品,则可能面临 AI 馆员无人问津、资源浪费的局面。面对这一问题,图书馆首先需明确,智能化建设不仅仅是简单的技术堆砌,而是一项综合性的运营与管理工程,其推进工作需建立在充分的前期论证之上,精准分析本馆的业务需求、资源条件及发展定位,并在此基础上制定科学合理的实施策略,避免“一刀切”式的部署思路。在实践中,应强调高层决策的引领作用,同时充分调动全体馆员的积极性,在实际推进过程中努力锻造图书馆智能化应用的运营与管理能力,提升智能化服务的专业水平。此外,图书馆还应高度重视本地数据库、知识库及语料库的建设,这不仅是提升 DeepSeek 应用价值的关键,更是图书馆 DeepSeek 区别于市场竞品、发挥特色优势的重要支撑。通过加强本地知识库构建,结合特色馆藏资源数据,图书馆可有效提升 DeepSeek 的专业性、精准度与实用性,从而真正实现 AI 技术与图书馆业务的深度融合与长远发展。

2.3.2 算力资源问题

尽管 DeepSeek 在应用层面对算力需求相对较低,但在企业级应用场景下,仍需要强大的计算资源支撑。如被称为“满血版”的 DeepSeek-R1,其参数规模高达 6710 亿,最低算力要求为双 H100GPU 和 1TB

内存,而推荐配置则是由8张A100 80G GPU组成的服务器集群,并配备相应的硬件优化与管理维护能力。这一高昂的算力需求对于多数图书馆而言是一个重大挑战,因此,建立算力共享平台成为必要的解决方案^[30]。图书馆可以通过建设算力网络,将分布式计算节点互联,实现算力与网络的一体化提供、一体化调度和一体化交易,并针对不同类型的图书馆,采取差异化的算力共享策略。对于公共图书馆,可以依托总分馆模式,由总馆建设并运维专用的算力中心,统一调度和管理计算资源,从而有效弥合基层图书馆在算力条件上的差距。对于高校图书馆,在对DeepSeek进行本地化部署之前应充分考虑学校已有的算力基础设施,优先依托校级算力中心,以降低建设与维护成本。此外,根据实际经费条件,高校图书馆还可适当自建低算力硬件环境,主要用于学术研究的实验、测试等轻量级任务,确保在资源受限的情况下依然能够开展基础的AI应用探索。

2.3.3 用户使用风险

用户在使用图书馆DeepSeek的过程中可能会存在一定风险。DeepSeek大模型虽然在处理结构化信息任务上表现优异,但也存在一定的认知局限,在部分情况下可能会生成看似合理但实际错误的内容,此类“幻觉”问题在图书馆DeepSeek的参考咨询、知识问答、情报分析等场景中可能会直接影响用户的信任感和使用体验。同时,DeepSeek在图书馆领域的广泛应用可能会引发用户的“滥用”风险。部分高校图书馆构建的Deepseek科研支持服务场景虽然可以助力师生的科研工作,但也可能导致师生忽视独立思考、引用错误信息,甚至滥用AI辅助完成学术写作,影响科研诚信。面对此类风险问题,图书馆应尽快建立DeepSeek生成内容的审查与纠错机制,定期对系统的生成内容结果进行抽查与评估,确保其准确性与可靠性,并在发现问题时及时优化算法、调整权重,降低错误内容的出现概率。同时,图书馆还应实施全面的用户AI素养培训计划,培养用户的批判性思维,帮助其理解AI的局限性,提升对大模型生成内容的辨别能力,引导用户合理使用DeepSeek,避免盲目信赖。此外,为进一步规范用户行为,图书馆需建立科学合理的违规处理机制,实时

监控用户对于DeepSeek的使用行为数据,预防滥用风险,通过合理的引导与管理,确保AI技术在图书馆环境中的安全、高效与合规应用。

3 “图书馆+DeepSeek”的应用路径

DeepSeek浪潮对于我国图书馆界推动智能化转型是一次难得的良机,然而就像面对信息化、网络化、数字化的浪潮一样,图书馆需要在应用过程中不断探索才能开辟出可行的道路。对图书馆面对信息技术革新的探索与研究进行溯源时可以发现,其整体的发展路线可归纳为“战略提出-建设落地-运营管理-评价反馈”^[31-33]。首先,通过顶层战略架构设计明确认知与规划,为后续进程做好全局部署;其次,根据战略规划落实建设项目,统筹推进相关应用;再次,对图书馆的相关转型工作进行运营管理,保障工作的高效运行;最后,针对各环节的实际推进情况进行评估反馈,发现并改善可能存在的问题,进一步提升图书馆转型效率。基于此,本文提出以“战略层-建设层-运营管理层-评价层”为核心的“图书馆+DeepSeek”应用路径框架(见图1)。

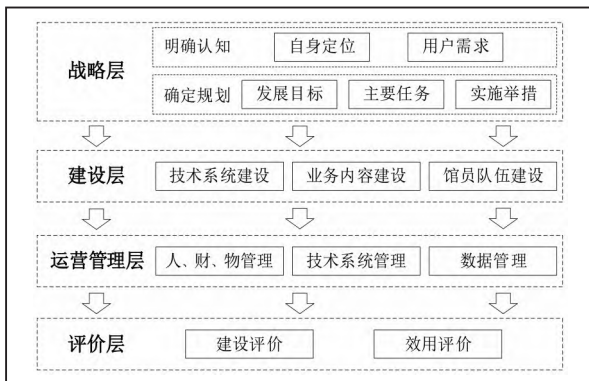


图1 “图书馆+DeepSeek”应用路径框架

3.1 战略层

图书馆应用DeepSeek实现智能化转型,需要在明确认知的基础上确定未来工作规划,构建合理、适宜的系统化战略框架。在认知层面上需做到明确自身定位与用户需求,在规划层面上需做到确定发展目标、主要任务及实施举措。

3.1.1 两个认知上的“明确”

(1)明确自身定位。一方面,图书馆的类型可根据服务对象的不同划分为公共图书馆、学校图书馆

以及专门图书馆^[34],同一类型图书馆之下又可细分出多种类型的图书馆,不同类型的图书馆在角色定位上的区别将影响 DeepSeek 的应用情况。如公共图书馆的角色定位偏向于“社会服务者”,那么其对于 DeepSeek 系列产品的选择应适用于服务社会大众群体,而高校图书馆的角色定位偏向于“师生服务者”,这种情况下的选择更需要考虑相关产品对于师生科研工作的支持水平等。另一方面,图书馆还需要明确自身的智能化成熟度水平,如通过评价方法明确本馆的智能化建设成熟度^[35],以此为依据选择适宜的 DeepSeek 产品对已有的智能化服务进行补充,从而做到资源分配的高效化。

(2)明确用户需求。图书馆的功能设计与服务安排需以用户需求为导向,DeepSeek 应用的选择与设计亦是如此。图书馆的用户类型大致可分为阅读型、设施型、活动型以及休闲型等^[36],其需求以文献资源服务、空间设施服务、文化教育服务、休闲娱乐服务以及信息咨询服务为主^[37]。图书馆应通过广泛调查明确本馆的用户需求,并以此为导向选择最能够提升用户满意度的 DeepSeek 产品,更有针对性地提升服务质量。

3.1.2 三个规划上的“确定”

(1)确定发展目标。图书馆对于自身定位及用户需求有了清晰的认知之后即可开始确定发展目标,为应用 DeepSeek 改善图书馆的服务质量指明方向。在建设目标层面,需确定 DeepSeek 技术系统和业务内容的建设目标,以及具备高数智素养的馆员人才队伍建设目标。在运营管理目标层面,需从人、财、物、技术、业务、安全管理等方面确定对应目标,如 DeepSeek 使用率、满意度指标,DeepSeek 结合图书馆本地知识库的垂直训练效果指标,相关的安全治理标准制定与执行目标等。此外,发展目标的制定还需具有阶段性,短期目标以完善图书馆 DeepSeek 的基础应用功能和提升使用率为主,中期目标以构建富有图书馆本地特色的垂直领域应用为主,长期目标则需聚焦构建基于 DeepSeek 的图书馆智慧服务生态^[37]。

(2)确定主要任务。如果将目标比作是价值导向的“终点”,那么任务就可以理解为行动导向的“路

径”。图书馆应秉持“目标牵引任务,任务支撑目标”的原则,在目标的指引下,统一规划 DeepSeek 应用的主要任务。在建设任务方面,图书馆需确定 DeepSeek 的本地化部署参数型号、DeepSeek 的应用功能内容,以及人才团队培养的建设任务等。在运营管理方面,图书馆需在对应目标下进行任务分解,如在 DeepSeek 的使用率目标下,确定相关的宣传推广、AI 通识教育、AI 实践活动等任务;在 DeepSeek 的满意度目标下,确定相关的满意度评估、定期更新 AI 系统等任务;在 DeepSeek 的垂直领域应用目标下,确定本地知识库、数据库的分阶段训练任务等。同时,图书馆还需在规划中确定本馆的重点任务与次要任务,确保馆员对于重点工作的明确,更有针对性地推动 DeepSeek 在图书馆的应用工作。

(3)确定实施举措。图书馆对于 DeepSeek 的应用工作应以发展目标和主要任务为指导,在分步实施中稳步推进。将目标与任务按年度、季度、月度为单位进行分解,以重点任务为牵引,逐步完成推进工作。图书馆也可采取“总规划、局部先行”的策略,从某些要素与环节切入,再推广到全要素、全流程,如先从 DeepSeek 如何应用于图书馆参考咨询服务切入^[38],在对相关的 DeepSeek 模型、图书馆数据库、知识库、用户交互、图书馆系统等要素进行布局并成功应用之后,再以此为经验基础逐步推广至图书馆的阅读服务、智库服务等。此外,从图书馆行业视角来看,图书馆还可以进行试点先行,通过先进示范引领更多的图书馆构建基于 DeepSeek 的数智化服务体系。

3.2 建设层

图书馆应遵照战略规划部署,依照自身的优势、特色等选择适合的建设路径与方案,形成富有特色的建设模式,统一推进“图书馆+DeepSeek”的技术系统建设、业务内容建设以及馆员队伍建设。

3.2.1 技术系统建设

图书馆推进 DeepSeek 大模型技术系统建设的路径主要可以分为两类,分别对应通用场景领域和垂直场景领域。一方面,图书馆可通过检索增强生成(Retrieval Augmented Generation, RAG)技术,直接调用 DeepSeek 开源大模型接口,结合本地构建的领域知识库,提供相关服务。目前大多数开始应用

DeepSeek的图书馆都是采用该建设路径,其面向的主要是以智能问答功能为代表的通用场景需求。RAG技术可通过获取外部知识库信息提升大模型的生成能力,结合知识库可动态更新的特点,模型生成内容的准确性和时效性可以得到有效提升。另一方面,图书馆还可通过构建特色语料库,结合微调参数,训练具有本馆特色的垂直领域大模型,如前文提及的常州市图书馆部署的常州地方文化大语言模型,而高校图书馆可通过聚焦学科特色,构建以高质量学术文献、项目报告为核心的特色语料库,训练针对特色学科的大模型。此外,图书馆在建设DeepSeek大模型技术系统时还应全面考虑本馆的硬件条件、经费预期、效果预期等因素,对大模型版本进行合理化选择。同时,图书馆界还应建立和优化数据交换与共享机制,利用技术手段消除各图书馆之间的数据共享障碍,实现跨馆的数据互联互通,推动单一的文献服务向全面的数据服务升级转型^[37]。

3.2.2 服务内容建设

基于DeepSeek的图书馆服务内容建设可参考前文提及的各行业对于DeepSeek的应用态势:一是推动DeepSeek与图书馆原有服务的进一步融合;二是对图书馆创新服务的有效实现。首先,图书馆服务的类型包括阅读服务(借阅服务、导航与检索服务、推荐服务、阅读推广等)、情报服务(科技查新与情报分析等)、知识产权信息服务、咨询服务、智库服务、学科服务、科研支持服务等。从目前已经开始应用DeepSeek的图书馆来看,大多数已实现DeepSeek与阅读服务、咨询服务的初步融合,能够支持用户通过接入DeepSeek的智能问答系统或AI馆员使用馆务查询、知识问答、读物推荐等服务。未来图书馆需要尝试进一步提升DeepSeek的业务应用广度与深度。在广度上,加速推进未融合业务模块与DeepSeek的整合进程,如将知识产权信息服务、情报服务、学科服务等融入DeepSeek系统,确保全方位覆盖与深度协同;在深度上,持续深化已整合业务模块与DeepSeek的系统耦合度,重点推进智能检索系统、云端书库、智能数字阅读平台等场景的底层架构重构,实现API趁热接入到智能中枢的深度嵌入,进一步提升用户体验。其次,基于生成式AI的场景创新已

逐渐成为近两年图书馆服务范式跃迁的战略支点,除了常州市图书馆推出的地方文化大语言模型等案例之外,图书馆还应积极挖掘潜在的应用场景,如支持沉浸式阅读的AI阅读助手、支持精准学术评价的AI科研助理、基于虚实融合的AI虚拟图书馆服务等,通过多维度的场景创新丰富图书馆应用DeepSeek的服务内容建设。

3.2.3 馆员队伍建设

面向DeepSeek应用的图书馆馆员队伍建设,应围绕“AI+专业馆员”的核心要点,构建能够适应AI时代的馆员团队。在具体操作上,图书馆可通过分层培养的方式,构建以基础层、中间层、顶层为构成要素的“金字塔型”能力结构馆员队伍。对于基础层,可采取AI素养教育的方式,开设AI通识教育、AI实践教学等培训课程或讲座,提升全体馆员的AI素养,如国防科技院校图书馆联盟于2024年5月开展的馆员AI培训会,为打造具备较高AI素养的馆员团队贡献力量^[39];对于中间层,可针对业务场景进行骨干力量的定制化培训,针对图书馆的AI相关核心业务进行馆员培训,如四川大学图书馆将馆员的AI培养划分为编目自动化组、科技查新组、学科分析组等八个小组,通过“项目化学习”提升图书馆各业务方向的AI人才储备^[40];对于顶层,图书馆应积极引进与联合培养DeepSeek相关的专家型人才,将“AI+图书馆学”纳入新的馆员招聘标准,并与高校、科研机构合作,联合培养交叉学科馆员,储备顶尖型馆员人才。

3.3 运营管理层

对于DeepSeek相关的建设成果,图书馆应以优化用户体验为导向,努力锻造其运营与管理能力,打造富有成效的运营与管理模式,具体可从“人、财、物”管理、技术系统管理、业务管理、安全管理等方面提升其智能化服务的专业水平。

3.3.1 人、财、物管理

“人”的管理是指对图书馆DeepSeek用户的管理,即图书馆对用户的信息、行为和需求进行管理与优化的过程,旨在为用户提供更高效、便捷、个性化的服务。首先,图书馆应对本馆的DeepSeek用户的数据信息进行系统化管理,如借助AI工具自动收集、分类和更新用户的注册信息和历史记录信息,并通过

分析用户信息建立用户画像,初步了解用户使用DeepSeek的相关需求。其次,图书馆应根据DeepSeek的平台系统日志采集用户的访问行为,在此基础上对其行为数据进行分析,提取用户行为模式,识别用户偏好,如国家图书馆通过大模型问答平台的日志分析提取出用户的不同需求^[41]。最后,在整合了用户使用DeepSeek的相关需求之后,还需通过实时数据分析监控用户需求变化趋势,针对不同需求提供差异化服务,为用户提供更加高效、便捷的服务体验。

“财”的管理即是对图书馆应用DeepSeek的经费管理。有研究表明,目前我国图书馆开展AI应用的经费投入还未形成较大规模,大多数图书馆的经费预算不足成为推动AI应用发展的主要困境之一^[42]。面对这一问题,图书馆应进一步细化管理开展DeepSeek服务的相关经费条目,统筹管理生产端技术系统的开发投入、消费端的软硬件投入和交互设备的购置投入等,在有限的经费条件下选择性价比最高的建设与运营方案。

“物”的管理是指图书馆应用DeepSeek的相关资源管理,即图书馆对DeepSeek相关的各类资源进行组织、更新和共享的全过程,从而达到资源的合理分配、设计和优化的目的。DeepSeek相关的资源组织主要聚焦于图书馆知识库与语料库数据的整理、分类与标注,通过特色知识库与语料库的构建与使用,充分发挥本馆的资源特色;资源更新是图书馆根据用户的需求变化、文化发展、社会趋势,对DeepSeek的相关资源进行补充、淘汰、替换和优化的动态管理过程,图书馆应对知识库即训练语料库进行实时更新,从而保障DeepSeek的准确性和时效性;资源共享是多个图书馆或信息服务机构通过合作,将各自的资源互联互通,共同为用户提供更广泛的DeepSeek服务的过程,该项工作可以参考基于区块链的联盟链平台(如FISCO BCOS联盟平台^[43])等方式实现,其高易用性、高性能、可运维性、高安全性等特点可以保障图书馆联盟的有效使用,图书馆可据此构建专属的联盟链平台,实现DeepSeek资源的高效流通,促进相关单位之间的资源共享。

3.3.2 技术系统管理

技术系统管理是指图书馆对DeepSeek相关的

各类系统进行技术创新改造、系统更新、系统维护的过程,从而进一步确保图书馆DeepSeek的安全、稳定、高效运营。首先,图书馆可利用新一代信息技术助推图书馆DeepSeek的创新改造,如将元宇宙技术和DeepSeek模型相结合,可让DeepSeek在虚拟空间以虚拟人的形式与用户进行交互,让DeepSeek根据用户的兴趣推荐资源或引导其探索相关区域,为读者提供全新的图书馆使用体验。其次,当前AI大模型行业发展态势迅猛,图书馆应积极关注包括DeepSeek在内的开源大模型的最新研发进展,通过及时跟踪该领域的技术动态,确保在更高性能和更优性价比的大模型出现时,能够实现及时接入,为用户提供更加便捷和高效的服务。此外,在图书馆DeepSeek系统的日常运行中,图书馆应充分利用AI技术进行运行质量的实时监控和自动化维护,如借助智能监控工具实时检测系统的健康状态,快速响应潜在的技术故障,并安排专员或采取自动化措施进行及时修复;还可通过自动化维护进行定期的性能优化和更新管理,确保DeepSeek始终处于最佳运行状态,减少系统停机时间,保障用户的持续访问和使用。

3.3.3 数据管理

数据是大模型的基础,是决定大模型应用效果的关键因素之一。图书馆需要制定相应的数据使用规范、标准和流程,以确保DeepSeek数据的质量、可靠性、安全性和正确性。DeepSeek相关的数据管理要以其应用的发展目标为宗旨制定相关的数据策略,既要考虑到DeepSeek应用的生命周期,如数据存档、加工复用、版本更新、质量检测、备份恢复等,还要重视法律、规定、版权、行业标准和道德等方面的要求,以确保应用的合规性^[44]。对于图书馆DeepSeek的数据而言,管理难度较大的是文献内容数据和支持图书馆系统的知识库数据^[37],前者需要在获得出版社和作者授权的基础上,通过专门的AI工具进行处理,其对技术、算力和存储等方面有一定要求;后者则需要制定中长期规划,分批次、分阶段实施,确保知识库的逐步完善与持续更新。此外,DeepSeek的本地化部署在一定程度上能保障数据的安全性,但与用户的智能交互过程中,可能会出现数据安全问

发一些敏感数据,面对此类情况则需要加强人工干预,对用户与图书馆 DeepSeek 的交互过程进行整体性的监督审核,保障数据安全^[38]。

3.4 评价层

针对 DeepSeek 相关的建设和运营管理成果,图书馆应致力构建科学、合理、可操作的图书馆 DeepSeek 评价体系,全面衡量 DeepSeek 相关的建设质量与实际效用,为图书馆优化相关工作提供反馈数据,具体可从“建设评价”和“效用评价”两个方面进行评估反馈。

3.4.1 建设评价

图书馆 DeepSeek 相关的建设评价可聚焦于技术系统建设评价、服务内容建设评价及人才队伍建设评价等方面,以确保 DeepSeek 能够高效融入图书馆的智能化发展进程。

在技术系统建设评价方面,需重点考察 DeepSeek 技术体系的合理性,评估其是否依据本馆的业务需求选择了合适的大模型产品,确保其在智能问答、知识检索、个性化推荐等方面的能力能够精准匹配用户需求。同时,系统的稳定性与安全性也是关键评价因素,需确保 DeepSeek 能够在高并发环境下稳定运行,并采用了严格的数据安全与隐私保护措施。此外,数据更新的及时性也是重要指标,需评估图书馆是否有按照既定周期对知识库进行数据更新,以确保 DeepSeek 输出信息的准确性、时效性。

在服务内容建设评价方面,需重点关注 DeepSeek 与图书馆现有服务体系的融合程度以及场景创新的挖掘和落地情况。如 DeepSeek 是否能够深度嵌入图书馆的阅读服务、情报服务、学科服务等传统业务,推动图书馆服务的智能化升级。同时,还需评估其在创新应用场景中的探索进展,确保其能够真正满足不同用户群体的需求,充分发挥大模型的应用潜力。

在人才队伍建设评价方面,图书馆应积极对开展的 DeepSeek 培训内容进行系统性评估,确保馆员能够熟练了解培训中相关技术并应用于实际服务之中。同时,还需对人才招聘标准进行评价,确保图书馆能够有效引入符合 DeepSeek 背景的专业型人才。此外,还应尽快构建科学合理的馆员数智素养

指标体系,以量化评估馆员在 AI 技术环境下的业务能力、技术理解能力和服务能力,推动馆员队伍的全面发展,确保 DeepSeek 的技术潜力在专业队伍的引领下能够发挥最大化收益。

3.4.2 效用评价

图书馆在对 DeepSeek 进行效用评价时,应重点围绕与用户的交互效果评价、用户满意度评价以及具体应用场景的使用效果评价等方面展开,以确保 DeepSeek 在融入图书馆之后能够具有较好的运行效果。

在交互效果评价方面,图书馆需重点关注 DeepSeek 交互界面的易用性指标,尤其是面对不同用户群体时,交互界面的操作是否简便直观,确保用户可以快速上手并流畅使用。同时,还需要评估 DeepSeek 在与用户交互过程中的响应速度、生成内容的准确性和相关性,以及其在复杂检索任务中的智能化程度,确保其在具备输出效率的同时,还能提供高质量的内容支持。

在用户满意度评价方面,图书馆应积极提升用户评价的参与度,建立完善的用户反馈机制,可采用问卷调查、用户访谈、在线评分等方式收集用户反馈,综合评估 DeepSeek 在不同类型用户群体中的适用性,并根据反馈不断改进 DeepSeek 的系统功能,提升用户满意度。

在具体应用场景的使用效果评价方面,需要针对不同应用场景设定相应的评价指标。如在科研支持场景中,需评估 DeepSeek 在学术文献检索、研究综述生成、开题报告撰写以及学术伦理方面的表现;在阅读服务场景中,需评估 DeepSeek 在读物推荐的精准度、阅读服务功能的适用性以及其在用户个性化阅读体验中的实际效果;在学科服务场景中,则需评估 DeepSeek 的学科交流能力、学科文献搜集与推荐能力以及学科情报分析能力等。

4 结语

DeepSeek 带来的算法技术及市场策略突破,加速了大模型技术从实验室向实际应用的落地工作,为图书馆的智能化转型提供了强劲动力。作为知识传播的重要枢纽,图书馆近年来不断深化数字化变

革,推动智能化的知识服务体系建设,构建智能化信息管理与用户服务模型。从 ChatGPT 的爆火到如今 DeepSeek 的广泛应用,生成式 AI 正逐步融入图书馆的各类服务场景,推动图书馆行业的创新步伐不断向前。本文从 DeepSeek 的突破性特征分析出发,剖析了 DeepSeek 的应用态势与应用挑战,并搭

建了以“战略层-建设层-运营管理层-评价层”为核心的“图书馆+DeepSeek”应用路径。未来,图书馆界应继续审时度势,科学应对 AI 时代的多方面挑战,积极探索全新的智能化转型路径,发挥图书馆的精神内核,为用户创建更高效、更便捷、更智能的信息服务平台。

参考文献:

- [1] DeepSeek7 天突破 1 亿用户,成最快应用里程碑[EB/OL].[2025-03-01].https://www.sohu.com/a/856916586_122014422.
- [2] 张炜,高馨.公共图书馆 AI 应用现状与发展策略研究[J].国家图书馆学刊,2023,32(5):75-83.
- [3] 国家图书馆.嗨,很高兴认识你!我是“顾朝夕”[EB/OL].[2025-03-01].https://mp.weixin.qq.com/s/wRHytFpTz_v37yfx5bXnWQ.
- [4] ChatGPT 需要 1 万张 NVIDIA A100 显卡成本超 10 亿国内 6 家公司负担得起[EB/OL].[2025-03-04].https://www.toutiao.com/article/7218756912551199284/?upstream_biz=doubao&source=m_redirect.
- [5] Liu A, Feng B, Xue B, et al. Deepseek-v3 technical report[J]. arXiv preprint arXiv:2412.19437, 2024.
- [6] 张慧敏. DeepSeek-R1 是怎样炼成的?[J]. 深圳大学学报(理工版), 2025, 42(2):226-232.
- [7] Gloeckle F, Idrissi B Y, Rozière B, et al. Better & faster large language models via multi-token prediction[J]. arXiv preprint arXiv:2404.19737, 2024.
- [8] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017: 30.
- [9] 蔡天琪,蔡恒进. DeepSeek 的技术创新与生成式 AI 的能力上限[J]. 新疆师范大学学报(哲学社会科学版), 2025, 46(4): 136-143.
- [10] 李国杰. DeepSeek 引发的 AI 发展路径思考[J]. 科技导报, 2025, 43(3):14-19.
- [11] Guo D, Yang D, Zhang H, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning[J]. arxiv preprint arxiv:2501.12948, 2025.
- [12] 模型 & 价格 | DeepSeek API Docs [EB/OL].[2025-03-07].https://api-docs.deepseek.com/zh-cn/quick_start/pricing.
- [13] 最高优惠 75%! DeepSeek 推错峰优惠, 下调 API 夜间空闲时段价格[EB/OL].[2025-03-07].<https://baijiahao.baidu.com/s?id=1825130485410578794&wfr=spider&for=pc>.
- [14] DeepSeek · GitHub[EB/OL].[2025-03-07].<https://github.com/deepseek-ai>.
- [15] 周鸿祎谈 DeepSeek: 开启 AI 产业的“iPhone”时刻[EB/OL].[2025-03-07].<https://baijiahao.baidu.com/s?id=1823494973312054726&wfr=spider&for=pc>.
- [16] 深度开源助推 AI 大模型普惠“破圈”[EB/OL].[2025-03-07].<https://news.qq.com/rain/a/20250304A01GIL00>.
- [17] 阿里凌晨开源新模型: 性能对标 DeepSeek-R1, 参数仅为 1/20[EB/OL].[2025-03-07].<https://mp.weixin.qq.com/s/5MnuAu-ffKScHmWsCYJT8Q>.
- [18] 方兴东,王奔,钟祥铭. DeepSeek 时刻: 技术—传播—社会(TCS)框架与主流化鸿沟的跨越[J]. 新疆师范大学学报(哲学社会科学版), 2025, 46(4):126-135.
- [19] 全国率先! DeepSeek 丝滑入职广州 12345[EB/OL].[2025-03-10].<https://mp.weixin.qq.com/s/-QEYNl6n-5lrfi1FbVW7A>.
- [20] 人均 DeepSeek, 微博、知乎、微信被卷入 AI 搜索大战[EB/OL].[2025-03-10].<https://mp.weixin.qq.com/s/nip1G6rPfalhpafaOo6XLA>.
- [21] 一文梳理 98 家国资央企 DeepSeek 应用进展[EB/OL].[2025-03-10].<https://mp.weixin.qq.com/s/jEjqr1PFxNwSb5TGBcqBw>.
- [22] 部署 DeepSeek, 银行大模型矩阵扩容! 应用场景尚待深化[EB/OL].[2025-03-10].<https://mp.weixin.qq.com/s/qDj3Hrft>

- EMYY1Maowc6ysA.
- [23] 南通一院探索“AI+医疗”,深度对接 DeepSeek 效率 up up! [EB/OL]. [2025-03-10]. <https://mp.weixin.qq.com/s/b3VrFYr1ejtTKxolDJuWYyA>.
- [24] DeepSeek × 长江设计,真的来啦! [EB/OL]. [2025-03-10]. https://mp.weixin.qq.com/s/bGEuqKtk_OJMPqhQkaoSBQ.
- [25] 扬大 AI 馆员接入 DeepSeek 满血版:智慧服务重磅升级! [EB/OL]. [2025-03-12]. <https://mp.weixin.qq.com/s/zbDbvYaDrJuCuIYVNeHQ7Q>.
- [26] 兰州大学 AI 馆员正式接入 DeepSeek,开启智慧图书馆新篇章! [EB/OL]. [2025-03-12]. https://mp.weixin.qq.com/s/GZP1f-yGFrpvLSvh_iil2g.
- [27] 图书馆 AI 馆员×DeepSeek:智慧服务重磅升级 [EB/OL]. [2025-03-12]. https://mp.weixin.qq.com/s/GZP1f-yGFrpvLSvh_iil2g.
- [28] 学习地图全点亮! AI 馆员+DeepSeek 满血版重磅推出! [EB/OL]. [2025-03-12]. https://mp.weixin.qq.com/s/P9KFJvh0lr7wMQU_1Uh8Q.
- [29] 常州图书馆率先发布地方文化大语言模型 CZLIB 2.0 [EB/OL]. [2025-03-12]. <https://mp.weixin.qq.com/s/jEjqr1PFXnWdSb5TGBcqBw>.
- [30] 杨新涯,戴立伟,钱国富. DeepSeek 在图书馆的应用场景构架研究 [J/OL]. 图书馆论坛, 1-8 [2025-03-15]. <http://kns.cnki.net/kcms/detail/44.1306.g2.20250218.1639.004.html>.
- [31] 柯平,胡娟,袁珍珍. 新中国成立 70 年我国图书馆事业变革:路径、经验与问题 [J]. 国家图书馆学刊, 2019, 28(5):41-51.
- [32] 刘敬仪,刘静羽,初景利. 高校图书馆高质量发展总体设计与发展路径的思考 [J]. 大学图书馆学报, 2023, 41(4):22-28.
- [33] 邵波,许苗苗,王怡. 数据驱动视野下高校智慧图书馆建设及服务规划——兼论“十四五”时期智慧图书馆发展路径 [J]. 图书情报工作, 2021, 65(1):41-46.
- [34] 鲜鹏. 论图书馆类型划分的本质属性——兼论党校图书馆应该向研究型专门图书馆发展 [J]. 情报资料工作, 2002(6):48-50, 54.
- [35] 段美珍,刘惠,刘琪,等. 智慧图书馆建设成熟度评价方法及实证研究 [J]. 现代情报, 2023, 43(5):66-78.
- [36] 孟高慧,何家瑜,刘畅. 基于多元服务需求的图书馆用户类型划分研究 [J]. 图书情报知识, 2024, 41(6):41-53.
- [37] 刘畅,孟高慧,林博,等. 基于大规模用户调查的图书馆多元服务需求和满意度研究 [J]. 图书情报知识, 2024, 41(6):28-40, 53.
- [38] 罗燚,李广通,何杰. DeepSeek 赋能图书馆参考咨询服务的路径 [J/OL]. 图书馆论坛, 1-9 [2025-03-16]. <http://kns.cnki.net/kcms/detail/44.1306.G2.20250226.1248.003.html>.
- [39] 图书馆举办第十一届校(院)读书月系列活动——AI 应用培训 [EB/OL]. [2025-03-17]. <https://mp.weixin.qq.com/s/PYfmd0DrGjYHIrMenyDuw>.
- [40] 落实人才强馆战略 AI 赋能专业馆员——图书馆举行 2024 年“AI+业务”馆员培训工作总结会 [EB/OL]. [2025-03-17]. <https://lib.scu.edu.cn/node/128582>.
- [41] 翟蓉. 大模型赋能的智能问答 FAQ 语料库建设实践与思考——以国家图书馆为例 [J]. 四川图书馆学报, 2025(2):80-87.
- [42] 张伟,高馨. 公共图书馆人工智能应用现状与发展策略研究 [J]. 国家图书馆学刊, 2023, 32(5):75-83.
- [43] FISCO 金链盟简介 [EB/OL]. [2025-03-18]. https://www.fisco.com.cn/about_2.html.
- [44] 刘倩倩,刘圣婴,刘炜. 图书情报领域大模型的应用模式和数据治理 [J]. 图书馆杂志, 2023, 42(12):22-35.

作者简介:吴若航,南京农业大学信息管理学院博士研究生,研究方向:图书馆服务、智慧图书馆;茆意宏,南京农业大学信息管理学院教授,博士生导师,研究方向:信息行为、信息服务。