

# DeepSeek 推动下生成式 AI 走势及其图书馆应用前景探析\*

薛 霏<sup>1</sup> 王静静<sup>2</sup> 叶 鹰<sup>3, 4</sup> (1 浙江大学图书馆 2 山东大学新闻传播学院  
3 复旦大学国家智能评价与治理实验基地 4 南京大学江苏省国际联合信息学实验室)

**摘 要** 本文以 DeepSeek 为支点, 在梳理 DeepSeek 技术突破基础上阐明生成式人工智能 (Generative AI、生成式 AI) 的技术走势, 分析原有生成式 AI 格局的颠覆性重组。聚焦图书馆行业, 考虑从智慧图书馆变革升级、图书馆的个性化数字人文服务改善生活、图书馆获得差异化发展技术支持 3 个维度展开 AI 应用模型。揭示生成式 AI 驱动智慧图书馆变革, 并提出 AI 时代图书馆差异化发展的圆融发展方略。最后基于技术成熟度曲线预测未来 3–5 年的发展前景。

**关键词** 生成式 AI DeepSeek 智慧图书馆 数字人文 知识服务

DOI: 10.13663/j.cnki.lj.2025.05.004

## On the Trends of GenAI Driven by DeepSeek with Application Prospects in Libraries

Xue Fei<sup>1</sup>, Wang Jingjing<sup>2</sup>, Ye Ying<sup>3, 4</sup> (1 Zhejiang University Library; 2 School of Journalism and Communication, Shandong University; 3 National Experimental Base for Intelligent Evaluation and Governance of Fudan University; 4 Jiangsu International Joint Informatics Laboratory of Nanjing University)

**Abstract** The paper systematically examines generative AI (GenAI) developments based on the technological breakthroughs of DeepSeek, and analyzes the disruptive restructuring of the generative AI landscape. Focusing on the library applications, the study consider GenAI's potential across three dimensions: the transformation and upgrading of smart libraries, the personalized digital humanities services to improve life, and the access to differentiated technical support for library development. It also reveals the transformation of smart libraries driven by generative AI, and proposes a harmonious development strategy for the differentiated development of libraries in the AI era. Finally, it predicts development trends over the next 3 to 5 years based on the Gartner Hype Curve.

**Keywords** GenAI, DeepSeek, Smart libraries, Digital humanities, Knowledge service

### 0 引言

生成式人工智能 (Generative AI、生成式 AI) 的技术演进正经历从“规模驱动”到“价值驱动”的范式转换。随着 Transformer 架构

的持续优化<sup>[1]</sup>与训练范式的突破性创新 (如人类反馈强化学习), 生成式 AI 已从通用对话场景向知识密集型领域渗透<sup>[2]</sup>。在此进程中, 以 DeepSeek<sup>[3–5]</sup>为代表的中国技术力量通过知识

\* 本文系国家自然科学基金项目“基于多源异构数据的科技关键节点及信息扩散机理研究”(项目编号: 72304169)、山东省自然科学基金项目“科技‘种子’识别及信息扩散特征与方向研究”(项目编号: 23RWZX29)、山东省高等学校优秀青年创新团队项目“生成式人工智能背景下主流舆论建设及传播研究”(项目编号: 202RQ007)、山东大学中华民族现代文明建设研究专项“数字人文视域下黄河文化国际传播的影响因素、路径及提升策略研究”(项目编号: 23BQ086)的研究成果之一, 并受山东大学青年学者未来计划的支持。

通信作者: 王静静, E-mail: jingjingwang@sdu.edu.cn

增强预训练 (Knowledge-Enhanced Pretraining) 与混合专家系统 (Mixture of Expert, MoE) 的融合创新, 推动生成式 AI 进入“价值对齐”新阶段, 为垂直领域应用奠定了技术基石。这一技术跃迁正在重塑知识服务生态<sup>[6]</sup>, 而图书馆作为人类知识体系的核心枢纽, 其转型路径研究具有重要理论价值与实践意义。

然而, 当前研究仍存在一些局限: (1) 现有成果多聚焦生成式 AI 的通用场景能力评估 (如 OpenAI GPT 系列), 缺乏对垂直领域技术适配性的系统分析<sup>[7]</sup>; (2) 图书馆领域的生成式 AI 应用研究呈碎片化特征, 尚未形成“技术-场景-价值”的完整框架<sup>[8]</sup>; (3) 技术风险防控多停留于原则性讨论, 缺乏可操作的治理方案。这些不足制约了生成式 AI 在知识服务领域的深度落地。

针对上述问题, 本文以新近 DeepSeek 技术为支点, 围绕 DeepSeek 的技术突破, 分析生成式 AI 的技术走势<sup>[9]</sup>, 在此基础上指出生成式 AI 将会引发智慧图书馆再次变革, 进而探讨知识服务领域的智能化再升级方案。本文旨在为图书馆智慧化转型提供 AI 方法论, 同时为生成式 AI 的场景化落地提供参考。

## 1 DeepSeek 技术突破

2024 年 12 月 26 日, 杭州深度求索 (Deep Seek) 人工智能基础技术研究有限公司发布 DeepSeek-v3<sup>[10]</sup> 模型并同步开源, 性能媲美其对标的 OpenAI 闭源模型 GPT-4o, 训练成本仅为 GPT-4 的 5%–10%。2025 年 1 月 20 日, DeepSeek-r1 悄然亮相, 1 天后引发热议, 3 天后破圈, 6 天后震撼全球科技股, 28 日除夕, 这一基于 DeepSeek-v3 训练、专为复杂推理任务而生的 DeepSeek-r1 开源模型<sup>[4]</sup> 正式发布, 在多项测试中完美赶超其对标的 GPT-o1 模型, 成为春节期间世界热点话题并以最快速度接入社会生活, 也为知识密集型场景应用奠定了质优价廉的技术基础。

DeepSeek 最重要的优势集中在以较低的成本实现较强的 AI 性能, 其关键技术突破包括 3 个方面。

### 1.1 知识增强型架构创新

以 ChatGPT 为典型代表的生成式 AI 发展模式, 呈现出显著的技术工具主义特征: 其核心方法论依赖参数的指数级增长与算力资源的过度堆砌, 试图通过超大规模预训练突破性能瓶颈。这种建模方式, 本质上是对工具理性思维范式的延续, 因缺乏因果推断能力易引发决策黑箱化等系统性风险。而 DeepSeek 通过领域知识与神经网络的深度融合, 构建知识引导的注意力机制, 实现在领域知识融合和因果推理方面的跃迁<sup>[11]</sup>。

在领域知识融合方面, 其创新架构通过知识引导的注意力机制重构了传统模型的语义理解范式<sup>[12]</sup>。具体而言, 模型并非单纯依赖海量语料中的统计规律, 而是将医学、生物、法律等垂直领域的专家知识图谱编码为可计算的符号逻辑单元, 并与神经网络的表征学习能力动态协同。基于因果科学理论框架的创新突破, 在知识建模层面实现了双重进化。其核心技术架构通过多粒度知识网络的工程化构建, 不仅系统化整合了垂直领域的专家认知框架, 更在复杂系统因果建模维度取得显著进展。该技术方案创新性地融合了现代因果推理的层次化方法论, 在医疗决策支持等专业场景的应用研究中, 实验数据显示: 通过医学本体论与诊疗路径的深度嵌入, 系统在疾病推断任务中的效能指标实现显著优化, 其中推理精度提升幅度达四分之一, 同时数据标注资源消耗缩减至原有模式的十分之一<sup>[13]</sup>。

### 1.2 推理认知引擎

DeepSeek 在推理认知引擎领域带来成本、精确度、延迟性、内存等方面的创新, 其核心突破体现在混合专家架构优化方面, 如 DeepSeek-v3<sup>[6]</sup> 采用 671B 参数规模的 MoE 架构, 通过稀疏激活机制使单次推理仅调用 37B 参数, 在保持与密集模型相当计算成本的同时, 实现知识容量指数级增长 (对比 Qwen-2.5-72B 等传统架构)。在多头潜在注意力 (Multi-Head Latent Attention, MLA) 机制方面, 通过低秩分解将 KV 缓存压缩至传统 MHA 机制的 5%–13%, 结合“解耦 RoPE”位置编码技术, 在处理长序列时降低显存占用达

80% 以上 (DeepSeek-v2), 显著减少分布式推理的硬件需求。在多阶段训练策略方面, 采用群体相对策略优化 (PRPO) 强化学习, 跳过传统监督微调与人工标注环节, 使 DeepSeek-r1-zero 在无标注数据场景下仍保持优异推理能力, 降低数据获取成本约 40%。

其次, 在推理效果方面, 在训练阶段引入多 token 联合优化目标函数, 使 DeepSeek-v3 在 MMLU (Massive Multitask Language Understanding)、GSM8K(Grade School Math 8K) 等基准测试中准确率提升 12.7%, 同时通过动态路由机制确保推理阶段无须增加计算负载。DeepSeek 突破传统思维链 (CoT) 的线性推理模式, 通过问题复述、知识调用、反思验证等多元路径构建超过传统模型的思考深度。另外, 结合动态量化技术, 在保持模型精度的前提下将内存占用降低 62%, 支持在单卡设备上部署 8B 参数模型 (如 iPad 端侧运行案例), 实现移动端推理延迟低于 500 ms。

最后, DeepSeek 借助潜在空间注意力压缩和动态模型压缩技术, 实现了内存优化。潜在空间注意力压缩机制通过构建低维隐向量空间, 将传统注意力计算复杂度从  $O(n^2)$  降至  $O(nk)(k \ll n)$ , DeepSeek-v2<sup>[5]</sup> 处理 4096 tokens 长文本时内存占用减少 73%。集成 4-bit 量化、结构化剪枝与知识蒸馏技术, 生成轻量化模型版本。实验表明, 8B 参数模型在保留原模型 92% 性能的前提下, 显存需求降低至 1/4, 支持在边缘设备实现实时推理。

但是, 不可否认的是, 虽然 DeepSeek 的推理认知能力有了较大提升, 但其距离成熟的持续学习能力仍存在较大差距, 幻觉问题的存在表明模型在处理某些任务时仍依赖记忆, 而非推理能力, 举一反三效果仍需提高。

### 1.3 持续学习系统

DeepSeek 开源性为其持续学习系统提供了便利性, DeepSeek 通过“基础模型-社区协同”的双层演进机制以及跨模态持续进化能力为持续学习奠定基础。核心层通过密集发布 MoE 架构 (如 DeepSeek-v3)、推理加速模块等基础模型升级。生态层建立开源社区协同开发协议, 允许开发者基于领域数据 (如医学

文献、法律条文) 进行参数定制, 实验数据显示行业专用模型的领域任务准确率的提升。在 DeepSeek 中, 可构建多模态增量学习管道, 支持从文本模型 (DeepSeek-r1)<sup>[14]</sup> 到视觉语言模型 (DeepSeek-VL2)<sup>[14]</sup> 的无缝升级, 参数共享率大大提升。

总体而言, DeepSeek 成功的最重要关键在于其以较低的训练和运行成本、实现了高效的 AI 性能。表 1 是 DeepSeek-r1 与 GPT-o1 的部分技术参量比较。

表 1 DeepSeek-r1 与 GPT-o1 的部分技术参量对比

技术参量	DeepSeek-r1	GPT-o1
硬件配制 (Hardware)	2078 H800 GPU	28500 H100 GPU
软件训练方法 (Soft training)	RL + Distill	DL+SFT
参数量级 (Parameter scale)	671B	1800B
开源 / 闭源 (Open/proprietary)	开源 (Open)	闭源 (Proprietary)
资源需求 (Resource demand)	较少 (Less)	较多 (More)

由此可见 DeepSeek-r1 与其对标的 GPT-o1 相比, 投入效益或性价比非常之高。依据 DeepSeek-v3 报告, 其完整的训练过程共计 278.8 万个 GPU 小时, 假设每个 H800 GPU 的租赁价格为每小时 2 美元, 总成本则为 557.6 万美元, 仅为 GPT-o1 成本的 3%-5%, 打破了以“大模型、大数据和大算力”为核心的生成式 AI 扩展定律。让生成式 AI 随手可及、处处可用, 为各行各业应用 AI 带来了曙光。

## 2 生成式 AI 技术走势

在 DeepSeek 推动下, 原有生成式 AI 格局正在进行颠覆性重组, 具体表现在 3 个方面。

### 2.1 OpenAI 主导的“大力出奇迹”开始被颠覆

虽然 GPT-o3、Grok-3 等仍在“大而全”的轨道上奔驰, 但 DeepSeek 已开始在“小而精”的场景中落地。

有观点认为, AI 性能的提升主要依赖于规

模的扩展<sup>[15]</sup>。OpenAI 及 xAI 等公司在模型训练过程中依赖超大规模算力,而 DeepSeek 则以约 5%–10% 的成本实现同等性能,从而验证了“参数效率”优先于“绝对规模”的可能性,为“小而精”的 AI 发展范式提供了实证支撑。

近期由李飞飞研究团队推出的 S1 模型<sup>[16]</sup>引发了学界关注。该研究采取极简化的训练扩展方案,在测试阶段采用独特优化策略,仅通过千量级任务样本的训练,其逻辑推理能力评测结果已超过 OpenAI GPT-o1 模型的基准表现。值得注意的是,虽然该系统目前尚未完全实现多专家模块的精准功能分离,但实验数据有效支撑了该技术范式在当前算法优化中的实践价值。这种开始由“大力出奇迹”转向“小而精”的模式,为提升模型推理能力提供了可验证的新思路。

## 2.2 不一定以多模态起步,文本起步或许是快捷赶超的路径

DeepSeek 通过引入 MLA 机制,在算法架构层面实现知识架构创新。该模型在自然语言处理领域展现出卓越的泛化能力,不仅能精准捕捉文本深层语义关联,还可通过对抗式训练机制生成符合逻辑规则的语言序列,而这又为文本快捷赶超提供了一定的路径基础。首先,不同于传统模型依赖封闭式参数存储,DeepSeek 采用分布式知识图谱体系,通过动态路由机制实现领域知识的按需加载,使基础模型在保持一定参数量的同时,促使知识覆盖度达到较高的水平。其次,文本模态被赋予“认知基座”的战略定位,动态路由机制有助于文本场景率专业化分工的实现,这种“功能解耦”使 8B 模型在特定场景达到千亿级通用模型的性能。文本优先路径为多模态延伸提供结构化基础,动态路由机制在文本阶段已优化硬件资源利用率,节省的算力可定向投入多模态关键模块。

文本模态的“轻量化深耕”可产生双重红利:(1)通过开放架构实现知识资源的分布式积累,规避了多模态数据获取的高成本瓶颈;(2)动态路由机制在文本场景打磨的专家分工范式,为后续跨模态协同奠定认知基础。这种“单点突破-体系扩展”的技术路线,也可为后来者

超越既有行业巨头提供了可复制的战略框架。

## 2.3 开源生态

DeepSeek 通过其开源核心模型(主要是 DeepSeek-r1 和之前的 DeepSeek-v3),显著降低了行业布局大模型的门槛,吸引了大量开发者和研究者的关注,形成了独特的社区影响力。DeepSeek 正以领军开源 AI 之势,成为 AI 领域的明星。

开源策略不仅推动了学术研究的多样化发展,还促进了 AI 技术在工业界的广泛应用。尤其是将更加注重经济效益与可持续性发展。在经济效益上,2025 年将成为生成式 AI 领域投资回报率的关键拐点,以 DeepSeek 的崛起为例,其发展过程也遵循这一经济型趋势,在不断拓展应用场景中为企业降本增效,推动收入增长与市场份额提升。在可持续性发展方面,生成式 AI 聚焦能源效率、环境保护及政策支持。数据中心通过技术创新、绿色能源整合完善可持续性战略,政策上各国积极推动 AI 与各行业融合并注重可持续发展。

## 3 图书馆应用场景

图书馆在国产 DeepSeek 推动下必将进行 AI 重构,智慧图书馆建设将再次面临变革升级。

### 3.1 智慧图书馆变革升级

在生成式 AI 与智能计算技术的双重驱动下,智慧图书馆正经历从服务范式到空间形态的系统性变革。这种变革以知识管理体系的重构为核心逻辑,通过“知识管理全面接入-虚实场景全面贯通-空间服务全面升级”的三维路径:在知识管理维度构建智能闭环,在服务供给层面打造虚实共生的认知场域,在空间形态上完成从物理实体到元宇宙生态的迭代演进。

#### (1) 知识管理全面接入

图书馆知识服务基座呈现智能化演进特征。澳门科技大学图书馆于 2023 年搭建的 AI 大模型统一网关具有开创性意义,其通过集成多种 GPT 模型构建的 McMust 系统<sup>[17]</sup>,实现了知识服务的全流程闭环。该图书馆基于此技术基座研发的 ChatLib 人书对话系统,将知识服务从文献检索升级为知识建构,形成“组织-发现-生成-推荐”四位一体的知识服务跃迁。



2025年DeepSeek模型的部署更标志着技术底座的基础设施化转型,推动教学、科研与服务体系的深度智慧化融合。

知识生产流程的智能化再造在业务操作层面取得突破性进展。广东省立中山图书馆“采编图灵 LibRootSystem”<sup>[18]</sup>通过神经网络与机器视觉技术构建智能闭环系统:新书到馆后,系统自动完成验收信息提取(感知)、编目数据生成(决策)、分拣指令下达(执行)等全流程作业,使采编效率产生数量级提升——上架时间从20个工作日压缩至10分钟内。这种数据智能驱动的模式革新,不仅实现业务流程的标准化重构,更催生出实时动态更新的知识管理体系。

而DeepSeek开源框架的部署则进一步推动知识生态的动态化协同。该模型通过持续学习用户行为数据与学术资源更新,构建起动态知识图谱。在服务端,系统可依据研究者画像自动推送学科前沿动态;在管理端,支持馆藏资源与用户需求的实时匹配分析。对外从参考咨询到个性化服务,对内从编目到管理,皆可在生成式AI支持下实现业务升级和效率提升,让信息管理真正变成知识管理<sup>[19]</sup>。

### (2) 虚实场景全面贯通

图书馆作为知识传播的核心载体,正通过生成式AI实现物理空间与数字空间的认知融合。嘉兴市图书馆的创新实践具有示范价值:其五感阅读疗愈空间<sup>[20]</sup>,融合嗅觉、触觉等多感官刺激与人机交互装置,构建出沉浸式认知场域。值得关注的是,其自主研发的“AI绘梦大师”<sup>[21]</sup>,采用AIGC技术完成文本语义到视觉意象的跨模态转换,在“AI绘画之旅”活动中实现读者创意具象化,该案例验证了生成式人工智能对知识可视化传播的赋能效应。

基于此,未来图书馆元宇宙<sup>[22]</sup>的建构将突破传统服务边界,通过生成式AI构建虚实共生的认知生态:物理空间的文献资源经数字化拆解后,在虚拟空间重组为可交互的知识节点,读者可通过数字分身实现跨时空的知识漫游。这种双向赋能的融合机制,既保持了纸质阅读的深度思考特性,又延展了数字技术的沉浸优势,实现虚实场景全面贯通,并有效促进

信息与知识传播<sup>[23]</sup>。

### (3) 空间服务全面升级

图书馆智能化设备的应用可追溯至早期自动化系统,并随着人工智能技术的迭代持续演进。以美国北卡州立大学亨特图书馆BookBot机器人系统为代表的智能仓储解决方案,通过高密度自动仓储技术构建的智能化图书仓储系统,可实现200万册文献的精准存取,单次作业耗时低于5分钟,空间利用率较传统书库提升9倍<sup>[24]</sup>。智能盘点机器人系统通过多传感器融合定位与RFID识别技术,不仅突破传统人工盘点模式在效率与准确性方面的局限,更在文献定位、错架识别、资产清点等核心业务环节展现出显著的技术优势<sup>[25]</sup>。

在空间服务维度,人工智能赋能的硬件设备正在重构图书馆服务范式。基于机器视觉的智能分拣设备可实现文献流转的自动化处理,增强现实交互终端通过空间感知技术提供沉浸式导航服务,这些智能硬件通过多模态交互构建起“智能体-物理空间-用户”三元协同的智慧服务生态系统,不仅拓展了传统图书馆的功能边界,更为构建虚实融合的知识服务空间奠定技术基础。值得关注的是,现有研究已初步验证智能图书馆系统与元宇宙技术的深度融合路径,通过数字孪生、扩展现实等技术构建三维智慧知识空间<sup>[26]</sup>,这标志着图书馆服务正从物理实体向虚实共生形态演进,为智慧图书馆的范式革新提供新的研究方向<sup>[27]</sup>。

### 3.2 图书馆以个性化数字人文服务改善生活

智慧图书馆更加融入图书馆人文关怀,个性化数字人文服务也能在生成式AI驱动下更加“轻而易举”。试问当DeepSeek创作出如下诗句时,谁能怀疑未来AI图书馆整合全人类知识的超级智能与人文理想:

五绝·书海观潮:比特潮初涨,光纤夜未央。银屏浮孤岛,解码见霞光。

书山赋:巍巍秘阁,莽莽缥緲。芸帙裂云而鸣玉,绡帙垂霄以蔽荒。秋堂烁萤,耀雪案而窥天禄;冬架凝霜,卷缥緲以覆玄章。墨香浮乎秘阁,字海汇于沧浪。万轴缥緲,奔雷碾破蠹尘月;千寻邺架,灵犀衔来漆简光。书涛若怒,荡开混沌心源;文阵如禅,坐定古今行

藏。登楼而眺，八荒入彀；临架以揖，六艺在掌。书山非岳，自有宇宙洪荒。

此外，智慧图书馆在智能理解方面表现超越。以AI荐书与AI解书系统为例，其基于深度语义理解的技术架构，能够精准把握用户需求，提供兼具学术价值与人文温度的知识服务。在特殊群体服务领域，智能交互技术已取得突破性进展，如加利福尼亚州圣安娜公共图书馆部署的“AStounD”对话机器人，通过情感计算与多轮对话技术，有效辅助自闭症患者及社交能力较弱的儿童发展社交和情感技能<sup>[28]</sup>。随着DeepSeek等开源AI底座的持续优化，这类智能服务将在图书馆及更广泛的社会生活场景中加速落地应用。

正是凭借在汉语诗歌和智能理解方面超越ChatGPT的卓越表现，DeepSeek让大众相信AI支持下的智慧图书馆必能以个性化数字人文改善生活。真正能与人交流的智能机器人也见雏形，并将很快应用于图书馆和社会生活。

### 3.3 图书馆获得差异化发展技术支持

在生成式AI支持下，各类图书馆将获得差异化发展可能。面向公众服务的公共图书馆可以加大发展数字人文和AI4AH（AI for Arts and Humanities）的力度，面向科研服务的科研图书馆则可强化AI4S（AI for Sciences）和AI4SS（AI for Social Sciences）的深度，而面向师生的高校图书馆则可支持AI4AH、AI4S和AI4SS全面发展，这一AI时代差异化圆融发展方略如图1所示。

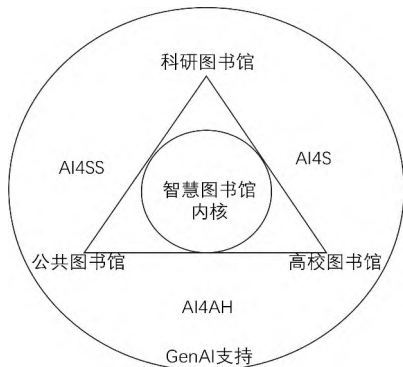


图1 生成式AI支持下的智慧图书馆差异化圆融发展方略示意

图1展示生成式AI支持下的图书馆发展可以维持共同的智慧图书馆核心，又依据服务对象的不同实现差异化发展，既考虑不同类型图书馆的功能定位，也兼顾AI4AH、AI4S、AI4SS的取向张力，构成AI时代的图书馆圆融发展方略。

## 4 未来发展趋势

依据技术成熟度曲线（Gartner Hype Cycle）<sup>[29]</sup>，生成式AI在图书馆领域的应用预期将于2026年进入实质生产高原期。重点发展方向预估有以下3个方面。

### 4.1 具身智能（Embodied AI）驱动的图书馆服务机器人

Embodied AI指机器人或者智能体能够通过感知、运动、理解和环境交互来获取信息、做出判断并采取行动的智能形式。Embodied AI驱动的图书馆服务机器人可通过先进的传感器和算法模拟生物的感知和认知能力，能够与复杂物理世界高效互动，例如，图书馆服务机器人可以利用计算机视觉技术识别图书条码，通过自然语言处理技术与读者交流，并通过深度学习优化其任务执行能力。此外，图书馆服务机器人还可以通过拟人化设计和多语言支持，进一步提升与用户的互动体验。随着技术的不断进步，图书馆服务机器人将朝着智能化、个性化方向发展，其发展不仅依赖于人工智能技术的进步，还需要结合图书馆管理系统的深度融合，以实现资源调度、数据分析等功能，推动图书馆服务模式的创新。

### 4.2 基于数字孪生的虚拟图书馆生态系统

数字孪生技术通过构建图书馆的虚拟模型，将物理空间、馆藏资源和用户行为等信息映射到虚拟空间中。这种技术的应用不仅能够实现图书馆运行状态的实时监控和资源需求预测，还能优化资源配置和服务策略。例如，管理人员可以通过数字孪生系统实时监控图书馆的运行状态，预测资源需求和用户流量，从而优化资源配置和服务策略。同时，读者也可以通过虚拟图书馆生态系统进行远程浏览和探索，提前了解图书馆的布局和资源情况。

此外，数字孪生技术在图书馆中的应用还

体现在其对智慧图书馆的赋能上。通过物理与虚拟空间的双向映射和交互,数字孪生技术实现了信息融合与服务创新。例如,数字孪生技术可以实现图书资源的数字化管理,提高图书检索效率,并通过虚拟仿真和数据分析支持空间布局优化和设施设备管理。此外,数字孪生技术还可以帮助图书馆实现智能化管理,降低运营成本,提升用户体验。例如,浙江大学未来图书馆计划用数字孪生技术构建虚拟阅读空间,通过3D可视化技术和边缘计算实现了图书馆资源的数字化建模和管理。此外,数字孪生技术还可以与元宇宙图书馆结合,为用户提供沉浸式、互动性强的虚拟阅读体验。

#### 4.3 跨机构知识网络的构建

通过图书馆联盟建设知识网络,让不同图书馆能够实现知识资源的共享与协同服务。各图书馆的知识资源无须离开本地,便可通过AI加密和安全技术实现跨机构的联合检索与利用,从而为用户提供更加全面的知识服务。这一方式不仅打破了信息孤岛,还将促进知识的流通与创新。

质优价廉的DeepSeek,必将在上述发展中展现强大动力和实用前景。

## 5 小结

DeepSeek作为自主研发的国产先进生成式AI代表,通过知识增强型架构创新、推理认知引擎和持续学习系统等在知识管理领域关键技术突破,必将给图书馆领域带来显著变革。在知识生产方面,DeepSeek通过高效的知识整合,将提升图书馆的知识资源建设效率;在用户服务上,其个性化推荐和智能交互功能将显著改善用户体验;在空间运营中,DeepSeek将助力图书馆实现智能化管理和资源优化配置。

固然,生成式AI技术在图书馆领域的应用也面临着诸多技术伦理挑战,包括知识权威性、用户隐私保护和数字鸿沟问题等。但挑战也是机遇,可以在确保技术应用安全性基础上努力维护可持续性,结合生成式AI在图书馆领域的应用将进入实质生产阶段,关注具身智能、数字孪生技术、跨机构知识网络以及个性化知识推荐与学习助手、多模态交互等发展方向。

生成式AI技术的发展和应用,可望为图书馆的数字化智慧化转型提供强有力支持,同时也为知识传播与服务创新开辟新的路径。

#### 参考文献

- [1] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. arXiv, 2017.
- [2] Bommasani R, Hudson D A, Adeli E, et al. On the opportunities and risks of foundation models[J]. arXiv preprint arXiv: 2108.07258, 2021.
- [3] Bi X, Chen D, Chen G, et al. Deepseekllm: Scaling open-source language models with longtermism[J]. arXiv preprint arXiv: 2401.02954, 2024.
- [4] 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.
- [5] Liu A, Feng B, Wang B, et al. DeepSeek-V2: A strong, economical, and efficient mixture-of-experts language model[J]. arXiv preprint arXiv: 2405.04434, 2024.
- [6] 张晓林. Library-Inside: AI赋能图书馆新质生产力的一种基础模型[J]. 中国图书馆学报, 2024, 50(3): 4-16.
- [7] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing systems, 2020, 33: 1877-1901.
- [8] Soudi M S, Bauters M. AI guidelines and ethical readiness inside SMEs: a review and recommendations[J]. Digital Society, 2024, 3(1): 3.
- [9] 叶鹰, 朱秀珠, 魏雪迎, 等. 从 ChatGPT 爆发到 GPT 技术革命的启示[J]. 情报理论与实践, 2023, 46(6): 33-37.
- [10] DeepSeek-AI. DeepSeek-V3 Technical Report [J/OL]. arXiv, <https://arxiv.org/abs/2412.19437v1>, 2024.
- [11] 令小雄. DeepSeek 开启后 ChatGPT 时代——基于数字范式革新及其运演哲思[J]. 西北工业大学学报(社会科学版), 2025(2): 59-67.
- [12] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in neural Information Processing Systems, 2020, 33: 1877-1901.
- [13] 程序员陆通. DeepSeek 为什么成本这么低[EB/OL]. (2025-02-08)[2025-03-06]. <https://baijiahao.baidu.com/s?id=1823498501696779670&wfr=spider&for=pc>.

- [14] Wu Z, Chen X, Pan Z, et al. Deepseek-vl2: Mixture-of-experts vision-language models for advanced multimodal understanding[J]. arXiv preprint arXiv: 2412.10302, 2024.
- [15] Sutton, Richard. The bitter lesson[J]. Incomplete Ideas ( blog ) 13, 2019( 1 ): 38.
- [16] Muennighoff N, Yang Z, Shi W, et al. s1: Simple test-time scaling[J]. arXiv preprint arXiv: 2501.19393, 2025.
- [17] 澳门科技大学. MeMust[OL]. [2025-03-15]. <https://itdo.must.edu.mo/page/WeMust.html>.
- [18] SCTECH. 标控科技. 广东省立中山图书馆“采编图灵”[OL]. [2025-03-15]. [http://www.sctech-gz.com/item\\_25625929\\_2787994.html](http://www.sctech-gz.com/item_25625929_2787994.html).
- [19] 叶鹰, 叶永鑫, 赵星. AI-GPT 技术及其应用生态下实现全面知识管理引论 [J]. 中国图书馆学报, 2024, 50( 6 ): 34-44.
- [20] 光明网. 浙江嘉兴市图书馆推出五感阅读疗愈空间 [OL]. [2025-03-15]. [https://culture.gmw.cn/2024-12/14/content\\_37738236.htm](https://culture.gmw.cn/2024-12/14/content_37738236.htm).
- [21] 朱文渊, 汤益飞, 邱仁博. AIGC 赋能阅读推广 [J]. 图书馆研究与工作, 2024( 11 ): 40-44; 57.
- [22] 赵星, 乔利利, 张家榕, 等. 元宇宙研究的理论原则与实用场景探讨 [J]. 中国图书馆学报, 2022, 48( 6 ): 6-15.
- [23] 王静静, 叶鹰. 生成式 AI 及其 GPT 类技术应用对信息管理与传播的变革探析 [J]. 中国图书馆学报, 2023, 49( 6 ): 15-24.
- [24] Kushins J. Let BookBot Bring You Any of This Library's Two Million Titles[OL]. [2025-03-15]. <https://gizmodo.com/letbookbot-bring-you-any-of-this-librarys-two-million-1541444173>.
- [25] 陆亚红, 马灵, 杨婉茹, 等. 智能图书盘点机器人技术的实践与应用研究 ——以上海交通大学图书馆为例 [J]. 大学图书馆学报, 2023, 41( 1 ): 44-51.
- [26] 周纲, 陈晨, 蔡文杰, 等. 虚实结合虚实相生的图书馆智慧空间服务 [J]. 图书馆建设, 2023( 4 ): 46-58.
- [27] 张慧, 佟彤, 叶鹰. AI 2.0 时代智慧图书馆的 GPT 技术驱动创新 [J]. 图书馆杂志, 2023, 42( 5 ): 4-8.
- [28] CSB News. California library uses robots to help kids with autism learn and connect[EB/OL]. [2025-03-15]. <https://www.wsgw.com/california-library-uses-robots-to-help-kids-with-autism-learn-and-connect/>.
- [29] Gartner Hype Cycle [EB/OL]. ( 2025-01-01 ) [2025-03-07]. [https://en.wikipedia.org/wiki/Gartner\\_hype\\_cycle](https://en.wikipedia.org/wiki/Gartner_hype_cycle).
- 薛 霏** 浙江大学图书馆, 馆员。研究方向: 数字图书馆、文献计量分析。作者贡献: 论文撰写及资料搜集。E-mail: [fxue@zju.edu.cn](mailto:fxue@zju.edu.cn) 浙江杭州 310058
- 王静静** 山东大学新闻传播学院, 副研究员。研究方向: 定量信息分析、智能信息传播。作者贡献: 论文撰写及修改、资料搜集。山东济南 250100
- 叶 鹰** 复旦大学国家智能评价与治理实验基地, 特聘研究员, 兼职资深教授。研究方向: 智能信息处理、定量信息分析。作者贡献: 论文选题、撰写及修改。上海 200433
- ( 收稿日期: 2025-03-10 修回日期: 2025-03-24 )