类 ChatGPT 人工智能技术嵌入智慧图书馆：应用价值、潜在风险及防控策略

刘凌宇[1]，徐中阳[2△]（1. 丽水学院图书与信息服务中心；2. 杭州电子科技大学管理学院）

摘　要：文章采用历史研究法、文献调查法以及理论分析法，在梳理 ChatGPT 的发展历程及其主要技术特征的基础上，深入分析了类 ChatGPT 人工智能技术嵌入智慧图书馆的应用价值、潜在风险以及风险防控策略。研究显示，类 ChatGPT 嵌入智慧图书馆具有提高资源建设质量、优化用户服务体验、实现人员数智赋能、促进管理决策转型等应用价值，但同时也存在知识产权保护难度上升、人机互动产生误导冲突、信息安全保护遭受冲击、管理决策过度依赖技术、复合型馆员有严重缺口等潜在风险。对此，智慧图书馆应制定构建版权协同保护体系、全面构建技术治理机制、完善数据安全保障体系、创新组织领导管理机制以及推动技术人力资源建设等风险防控策略，以增强风险防控能力。

关键词：ChatGPT；智慧图书馆；人工智能；应用价值；潜在风险；防控策略
中图分类号：G252；TP18　　**文献标志码：**A　　**文章编号：**1005－8214（2024）02－0045－11

Embedding ChatGPT-like Artificial Intelligence Technology into Smart Library: Application Value, Potential Risks, and Risk Mitigation Strategies

Liu Lingyu, Xu Zhongyang

Abstract: This article adopts historical research methods, literature surveys, and theoretical analysis to comprehensively analyze the application value, potential risks, and risk mitigation strategies of embedding ChatGPT-like artificial intelligence technology into smart libraries, based on an overview of the development history and key technical features of ChatGPT. The research indicates that embedding ChatGPT-like technologies into smart libraries has application values such as improving resource construction quality, optimizing user service experience, enabling personnel with digital intelligence, and facilitating management decision-making transformation. However, it also poses potential risks including increased difficulty in protecting intellectual property rights, misleading conflicts in human-machine interactions, impacts on information security protection, over-reliance on technology for management decisions, and serious gaps in composite librarian skills. To address these issues, smart libraries should implement risk mitigation strategies including establishing collaborative copyright protection systems, comprehensively constructing technical governance mechanisms, enhancing data security assurance systems, innovating organizational leadership and management mechanisms, and promoting the construction of technical human resources.

Keywords: ChatGPT; Smart Library; Artificial Intelligence; Application Value; Potential Risk; Risk Mitigation Strategy

1　引言

ChatGPT（Chat Generative Pre-trained Transformer）是由 OpenAI 公司研发的一种生成式人工智能语言模型[1]。该模型具备强大的自然语言理解能力和文本生成能力，能够实现与人类的多轮对话，且在交互中体现出形象人物化、功能语言化、理念人伦化、反馈精准化等特征[2-3]。令人惊讶的是，ChatGPT 的智能水平达到了史无前例

[基金项目] 本文系 2019 年国家社会科学基金重大项目“基于大数据的科教评价信息云平台构建和智能服务研究”（项目编号：19ZDA348）的研究成果。
△通信作者：徐中阳，hzhduxzy@126.com

的高度，ChatGPT 不仅能与人类进行智能聊天对话，而且在信息提取、文本写作、歌词创作等场景展现出极强的创造力[4]。因此，ChatGPT 自 2022 年 12 月发布以来便成为全世界关注的焦点，发布仅 2 个月便吸引了上亿的月活跃用户，成为史上用户增速最快的互联网应用。

ChatGPT 的突然火爆同样引起了学术界的广泛关注，国内外围绕 ChatGPT 的技术框架、运作模式、技术创新等内容，在教育教学、出版发行、体育科学等多个应用领域开展了深入研究并取得了大量成果。作为与数据、信息密切相关的学科，图书情报领域学者同样针对 ChatGPT 如何服务于智慧图书馆展开了深入的研究。储节旺[5]、李书宁[6]、吴若航[7]等学者深入分析了 ChatGPT 等人工智能工具为智慧图书馆服务带来的机遇、挑战，并在此基础上提出了有针对性的应对策略。蔡子凡[8]、郭亚军[9]、李立睿[10]等学者详细描绘了 ChatGPT 赋能图书馆智慧服务的具体内涵、主要特征、应用场景以及实施路径。

总体来看，ChatGPT 是人工智能领域的颠覆性成果，在各行业领域均有广泛的影响力和重要的应用价值。对于智慧图书馆而言，智能服务是其核心特征，而在 ChatGPT 上进一步开发得到的类 ChatGPT 技术作为真正意义上的人工智能，将进一步促进各项智能服务更加智慧化、智能化，为智慧图书馆的发展带来机遇与挑战。鉴于此，本研究在梳理 ChatGPT 的技术发展历程及主要技术特征的基础上，深入分析类 ChatGPT 嵌入智慧图书馆中的应用价值与潜在风险，并提出相应的风险防控策略，以期为未来智慧图书馆的建设提供理论参考。

2 ChatGPT 的发展历程及其主要技术特征

2.1 ChatGPT 的发展历程简析

由于 GPT 系列人工智能模型存在“黑盒”特性，因此本研究仅能通过对相关文献、公开报道、行业分析报告等资料对 ChatGPT 的发展历程进行梳理与归纳，得到表 1。由表 1 可得，ChatGPT 的发展经历了 5 个阶段。自 2018 年 6 月以来，OpenAI 公司开始推出 GPT 系列的生成式预训练语言模型。伴随着多次迭代，模型的参数量呈现出持续性的跨越式增长，由最初 GPT-1 的 1.17 亿个增长至 GPT-4 的约 3.5 万亿个。同时，随着多任务模型、大型语言模型、指令微调和基于人类反馈的强化学习等一系列关键技术在该模型中的应用，GPT 系列模型实现了大幅升级，更新至 ChatGPT，并在美国、英国、中国等多个国家掀起轩然大波，成为该系列中的代表性产品。此后，在 2023 年 3 月，GPT-4 问世。与 ChatGPT 相比，该模型具备更庞大的模型结构和更丰富的训练数据，进一步提升了自然语言理解、逻辑推理等交互应用能力。

表 1 ChatGPT 的发展

模型	发布时间	参数量	训练数据量	关键技术	技术对比
GPT-1	2018.06	1.17 亿个	5GB	无监督训练、有监督微调	具备一定的泛化能力，可接下游任务微调
GPT-2	2019.02	15 亿个	40GB	多任务模型	在包括聊天、续写、生成摘要等方面具备较突出的生成能力
GPT-3	2020.05	1750 亿个	570GB	大规模语言模型	实现对上下文小样本的学习能力强化，对于大多数 NLP 任务均可完成
ChatGPT	2022.11	1750 亿个	/	人类反馈强化学习方案	通过指令微调和人类反馈对话数据，实现强大的意图理解能力，反馈质量对齐人类喜好
GPT-4	2023.03	3.5 万亿个	/	基于规则的奖励模型	数据源得到进一步扩充，基于人类反馈调节，增加训练后处理和行为预测能力

2.2 ChatGPT 的主要技术特征

尽管 ChatGPT 是 OpenAI 针对聊天交互场景推出的智能聊天机器人应用，但其本质是对抓取信息进行处理并生成知识的全过程，与智慧图书馆知识服务需求高度契合。基于 ChatGPT 技术开发的智慧图书馆垂直领域类 ChatGPT 应用将在场景实现智慧服务提供技术支持。因此针对 ChatGPT 主要技术特征的深入分析与探讨同样具有重要意义。总体来看，ChatGPT 的主要技术特征可划分为 Tarnsformer、大型语言模型（Large Language Model，LLM）、基于人类反馈的强化学习（Reinforcement Learning with Human Feedback，RLHF）与指令微调（Instruction Tuning，IT）、多模态机

器学习（Multi-Modal Machine Learning，MMML）等四大类。

2.2.1 Transformer：自然语言处理能力的技术基础

Transformer 是 ChatGPT 获得自然语言处理能力的技术基础，其实质是一个基于注意力机制的深度神经网络模型。该模型的工作原理是通过引入自注意力机制，将不同的注意力权重赋予神经网络模型中的不同元素，进而根据元素的相关性对信息进行加权组合[11]。该模型主要包括编码器与解码器两部分，并应用下一单词预测（Next Word Prediction）、创新多头自注意力机制（Multi-head Self-attention Mechanism）来实现文字文本生成、上下文理解等自然语言处理能力。同时，结合对 GPT 系列模型的衍生版本的调查，在应用大小各异的参数量和微调后，Transformer 能够生成具备不同功能的模型，以适用于文本解析、关键词抽取、语言翻译、情感分析、复杂意图分析、特定受众文本摘要的生成等不同任务。可见，Transformer 的应用与创新为 ChatGPT 在智能咨询、交互聊天等多种应用场景下的文本生成提供了极其重要的技术支撑。

2.2.2 LLM：高质量文本输出能力的数据保障

大型语言模型（LLM）是一种基于深度学习的自然语言处理技术，具有极大规模的训练参数，同时它也是 ChatGPT 具备高质量文本输出能力的数据保障[12]。通过 GPT 模型迭代表现可以发现，高质量语料库资源是实现 GPT 自然语言处理效果的重要因素，且训练数据量越大，模型所能达到的自然语言处理能力越强。ChatGPT 所包含语料库大体可分为预训练语料与微调语料两部分。其中，前者包括维基百科、Gutenberg Book、Bibliotik Journey、Reddit links、Common Crawl 等，涵盖了在线百科全书、电子书、社交媒体及原始网页等内容，而后者则包括大量经过微调后的代码语料和对话语料。从功能上看，大型语言模型（LLM）通过对上述语料的深度学习实现对通用话语结构、语义及表达逻辑的认知，从而保障了高质量语言输出能力，另外由于语料库涉及内容宽泛，也使其具备了极强的世界知识能力。此外，大型语言模型（LLM）不仅提高了 Transformer 作用下各场景中的文本输出质量，而且为文本创作、高价值信息资源提取等知识创造、知识利用、知识传播场景提供了重要的语料基础，因而能够大幅度提高模型的知识价值。

2.2.3 RLHF 与 IT：人为介入机制的有效措施

基于人类反馈的强化学习（RLHF）是指利用强化学习方法，利用人类反馈信号直接优化语言模型。基于人类反馈的强化学习（RLHF）能使 ChatGPT 可以更好地理解人类所发出的指令，同时使输出结果更加真实和自然，且更符合人类偏好。值得注意的是，基于人类反馈的强化学习（RLHF）技术的实现需要奖励模型（Reward Model，RM）与近端策略优化模型（Proximal Policy Optimization，PPO）的共同作用，前者通过对生成答案进行打分排序，从而更新参数以强化生成答案的质量，后者则通过对 ChatGPT 进行微调与迭代，以实现文本生成策略的优化。指令微调（IT）是指在处理特定任务时对预训练模型进行调整以适应特定场景下的数据集和任务要求，如 ChatGPT 中的 Codex 模型，是以 GPT-3 为基础，通过大量编程代码、注释等数据微调得到的，使它在编程场景下具备极强的代码生成能力。总体来看，基于人类反馈的强化学习（RLHF）和指令微调（IT）的实质都是采用人为介入的方式进行原模型的调整和优化，通过人为介入机制，能够实现模型在伦理规则、法律政策等方面的引导，使其反馈效果越来越趋近于人类偏好，最终形成适配于不同垂直领域场景的衍生应用产品。

2.2.4 MMML：高质量交互场景的重要保证

多模态机器学习（MMML）是指通过机器学习的方法来实现处理和理解语音、文字、视频等多模态信息的算法，同样也是 ChatGPT 的主要技术特征之一。2023 年 3 月，OpenAI 发布了应用多模态机器学习技术（MMML）的 ChatGPT，即 GPT-4 模型。相较于前一代模型仅适用于语言模型任务的局限性，GPT-4 能够同时实现对图像、文字、声频、视频等多种模态信息进行知识提取、

逻辑推理、内容生成以及修改编辑等多种复杂操作，甚至在部分专业领域或学术科研领域也能表现出与人类相同的水平，极大地提升了模型结果的可靠性和准确性以及模型自身的通用性和泛化能力。可见，多模态机器学习（MMML）的应用使得 ChatGPT 更加趋近于成为一种人工智能通用大模型。另一方面，通过对更大规模的多模态信息资源的深度学习，ChatGPT 的用户需求判断能力、精准回复能力以及知识创造能力将会得到进一步的显著提升，从而能够生成完美契合用户需求的文本、图像、声频等不同类型的内容，为用户提供个性化、高品质、沉浸式等更多高质量的交互场景。

3 类 ChatGPT 嵌入智慧图书馆的应用价值分析

在 ChatGPT 掀起新一代人工智能革命后，诸如“文心一言”“通义千问”“LLaMa”等由各大科技公司开发的类 ChatGPT 产品先后问世，探索了更加丰富的应用场景，为更多垂直领域提供价值与可能性，而智慧图书馆场景成为类 ChatGPT 技术价值的集中体现。从应用价值上看，将其嵌入智慧图书馆中，能够显著提高资源建设质量、优化用户服务体验、实现人员数智赋能以及促进管理决策转型，推动智慧图书馆高质量发展。

3.1 智慧资源：提高资源建设质量

类 ChatGPT 嵌入智慧图书馆后能为图书馆的资源建设提供强大的技术支撑，进一步促进智慧图书馆中纸质文献资源、数字资源、知识资源等多维数据资源的深度融合，进而形成智慧资源。

首先，在文献资源建设上，类 ChatGPT 的应用将使资源采购更加科学。当前智慧图书馆常用的资源采购方式为读者荐购和 PDA 荐购，但二者均存在一定问题。前者存在反馈延迟、事后驱动、信息不完整等问题，后者则存在馆藏结构不合理、经费不足等问题[8]。而类 ChatGPT 能够依托自身采集的读者荐购数据、馆藏使用数据等大数据资源，客观分析读者阅读需求，并结合包括馆藏发展目标在内的决策指导数据，生成采购需求，最终据此在书目信息数据中生成推荐书目清单，实现馆藏文献资源的精准采购[9]。

其次，在数字资源建设上，类 ChatGPT 的应用能够优化资源结构、扩大资源体量以及促进资源融合。在数字资源结构与体量上，类 ChatGPT 一方面可以运用与上述文献资源采购相同的技术方式为图书馆提供客观、科学的商用数据库采购决策，另一方面可以通过对网络大数据的分析和抓取，将网络开放获取资源完整融入图书馆数字资源体系中，实现资源聚合，扩大数字资源内容，完善数字资源体系[13]。在促进资源融合上，类 ChatGPT 能够通过强化学习建立知识链接、知识结构、知识框架，实现对互联网中资源与馆藏数字资源的深度融合，构建一站式的资源获取平台，进一步强化用户与智慧图书馆之间的人机交互和资源利用。

最后，在知识资源建设上，类 ChatGPT 能够凭借其大规模、多模态数据的集聚与生产能力成为智慧图书馆知识资源建设的新引擎[14]。应用类 ChatGPT 技术的智慧图书馆，能够借助强化学习算法对馆藏资源进行文本挖掘和聚类整合，从而在显性知识分解、重构的过程中实现隐性化知识的归纳、保存与传播。同时，生成的隐性知识又将被智慧图书馆进一步显性化，成为新的馆藏知识资源。通过循环往复的资源交互、聚合分析，智慧图书馆将不间断地进行知识创造，其馆藏资源也将不断增值，使智慧图书馆真正成为“数智时代”的知识中心。

3.2 智慧服务：优化用户服务体验

身处数智时代，智慧图书馆用户对服务品质的要求日益增长，其主要体现在公共文化服务和学术科研服务两方面。类 ChatGPT 能为智慧图书馆提供技术支撑，借助 AI 赋能优化用户服务体验。

在公共文化服务上，类 ChatGPT 能为用户提供更加均等化、低成本、便利化的服务。依托类 ChatGPT 的智能技术，智慧图书馆与用户之间仅需通过自然语言交流，便可有效识别用户偏好及服务需求，并对其进行精准反馈。实现服务供给与用户需求的双向交互，降低公众

的使用门槛，从而降低用户在与智慧图书馆互动过程中所产生的学习成本、服从成本和心理成本，减轻用户使用公共文化服务过程中的行政负担。这样的服务模式可以有效弥合数字鸿沟，充分保障老年人、未成年人、残疾人等特殊群体的文化权益。

在学术科研服务上，类 ChatGPT 能为用户提供更加专业化、个性化的服务。① 优化科研辅助服务。借助类 ChatGPT 技术，智慧图书馆能为研究人员提供数据采集与保存、文本信息提取与生成、研究前沿分析与决策等一系列科研辅助服务，甚至能通过数据分析为科研人员提供科学选题、研究设计咨询等思维层面的引导[15]。② 强化科研定制服务。智慧图书馆能够充分发挥类 ChatGPT 技术在知识与技术方面的支撑作用，从而强化各类高质量的科研定制服务。例如，快速完成科研流程图、技术架构图、数据分析图等多种类型图表的构建与整合工作，进一步筑牢智慧图书馆建设中科研情报服务中心的重要地位。

3.3 智慧人员：实现人员数智赋能

身处智慧图书馆场景下，馆员同样是图书馆服务提质增效的重要一环。类 ChatGPT 嵌入智慧图书馆后，能进一步促进环境的高度智慧化，并实现对馆内人员的数智赋能。在类 ChatGPT 技术赋能下，馆内人员架构将会迎来较大幅度的调整，逐渐形成由智慧用户、智慧馆员以及虚拟数字人等三方人员构成的智慧人员，并通过共同作用不断完善智慧图书馆服务体系。

智慧用户是指将新一代人工智能技术应用于智慧图书馆之后，图书馆用户所获得的全新身份。类 ChatGPT 技术应用于智慧图书馆后，其智慧化的实现依赖于对馆内各类大数据的挖掘与分析，其中用户信息行为数据是保证智慧图书馆提供个性化服务及管理决策动态调整的关键内容。在用户满足自身知识需求的过程中，对用户信息行为数据进行追踪与管理，生成用户信息行为数据库，并基于此不断生成新的知识，同时通过反复的强化学习，完善图书馆各项服务与人员管理。可见，在智慧图书馆中，类 ChatGPT 能实现对人员的数智赋能，有效促进用户由普通用户向智慧用户的转变，使用户获得知识生产者和知识消费者的双重身份。

智慧馆员是指馆员为了适应智慧图书馆全新服务场景而形成的特殊角色，通常需要具备交叉学科背景和扎实的学科基础知识，不仅能基于用户需求提供专业的知识服务，同时也能胜任各类用户智慧素养提升工作[16-18]。类 ChatGPT 技术应用于智慧图书馆后，参考咨询馆员、学科馆员、数据馆员等类型馆员工作将逐渐被取代，而针对智慧环境下管理与服务的智慧馆员将大量增加。从岗位职责上看，智慧馆员一方面要充当类 ChatGPT 生成信息的审核与监管者，以保证交互过程中智慧服务的精准化、专业化和个性化，另一方面需要及时对交互过程产生的不当信息进行纠正，并通过标注对智慧图书馆类 ChatGPT 系统进行强化学习，进一步完善智能应用。

虚拟数字人则是指元宇宙环境下的智慧图书馆中存在的重要服务载体。在虚拟空间的智慧服务中，用户将直接通过虚拟数字人感受智慧服务。而在类 ChatGPT 技术的加持下，虚拟数字人将突破技术瓶颈，实现高度智能化，可以胜任虚拟讲解服务、阅读推广服务、用户培训服务等多种馆员工作，大幅提升智慧图书馆虚实交互下的用户沉浸式服务体验。同时，在虚实共生的智慧图书馆形态中，虚拟数字人能打破用户利用图书馆的地域限制和时间限制，通过智能交互引导用户虚拟分身链接至任意图书馆的目标资源，实现 24 小时不间断的服务供给，以满足虚拟环境下用户信息服务需求的激增[19]。

3.4 智慧管理：促进管理决策转型

类 ChatGPT 技术嵌入智慧图书馆后，能为促进管理模式转型、提升管理决策效能提供有力的技术支撑，进一步推动 AI 赋能下智慧图书馆的可持续发展。

在促进管理模式转型上，类 ChatGPT 的应用能促进智慧图书馆内管理的扁平化发展。类 ChatGPT 能够通过对智慧图书馆内各服务应用数

据的深度挖掘、建模分析等操作完成对服务需求的判断和反馈，继而以命令的形式传递至馆内的数字设备，实现对各类基础性工作的智慧化替代。与此同时，为了适应新的智慧化环境，图书馆将通过减少馆员规模、合并部门职能、精简传统部门等方式，应对其在组织规模、组织形态、组织机制等方面受到的影响，最终实现组织结构的扁平化发展[20]，有效降低部门间的协同沟通成本、提高主要业务的运行效率[21]。

在提升管理决策效能上，类 ChatGPT 技术能大幅度提升智慧图书馆管理层的战略规划决策能力。类 ChatGPT 应用于智慧图书馆后能实现对海量、多源、异构且类型多样的大数据资源的采集、存储、整理、分析以及再创造，推动语料库不断增值，为智慧图书馆管理决策提供准确及时的高质量数据支持。由于类ChatGPT 的相关决策与方案均基于智慧图书馆日常的智能采集数据，因而能使管理决策的依据更加客观科学，能避免传统管理中依赖个人经验作出主观决策的负面影响，为提高日常管理决策效能提供重要保障。此外，类 ChatGPT 还能对馆内空间信息进行实时监控并结合智能系统作出及时调整，实现对智慧图书馆操作性决策的智能化代理，使图书馆内温度、湿度及光亮程度等物理环境要素实现智能调控[22]。

4 类 ChatGPT 嵌入智慧图书馆的潜在风险分析

4.1 版权问题：知识产权保护难度上升

类 ChatGPT 嵌入智慧图书馆后，知识服务过程中所采用的数据与生成的知识产品是否具有可版权性，目前尚存争议。有学者认为类 ChatGPT 是作为内容生产工具进行内容的生成，其创作代表了程序设计者或训练者的主观意志，符合《中华人民共和国著作权法》的相关规定，是具有可版权性的[23-24]。基于此，本研究认为其所存在的版权问题表现在如下方面。首先，在数据的采集与挖掘上，为构建智慧图书馆垂直领域的类 ChatGPT 应用场景，其需要对馆藏资源及网络资源进行大量采集与挖掘，以达到智慧服务的预期效果。如类 ChatGPT 系统在未经允许的情况下抓取并使用受保护数据，将会严重侵犯版权方的复制权、信息网络传播权等著作权。其次，在内容的生成与使用上，类 ChatGPT 嵌入智慧图书馆后，在与用户交互中产生的反馈内容可能包括经过数据处理后生成的文献综述、学科前沿分析等高价值知识产品，但这些人工智能生成内容的版权究竟是归属用户、程序设计者还是智慧图书馆本身，将极易产生版权争端。当用户将智慧图书馆生成的知识产品进行复制、改编、传播时，也会产生侵犯复制权、署名权、修改权、信息网络传播权等著作权的风险。所以版权问题将存在于类 ChatGPT 技术处理数据的整个周期中，使知识产权保护难度上升。

4.2 技术偏见：人机互动产生误导冲突

尽管类 ChatGPT 能够表现出高度的智慧化特征，可以理解并分析自然语言的内容与情感并给予高质量回复，但智能程序是由代码生成的，仍然是人为创造的产物，即便经过反复的深度学习以及语料库的不断扩充，其生成的内容仍然在一定程度上体现出程序开发者的主观意愿，这也导致了类 ChatGPT 无法保证输出内容的客观、公正，换而言之，类 ChatGPT 存在一定程度且难以消除的技术偏见。而当类 Chat-GPT 应用在智慧图书馆场景中时，算法偏见、算法歧视等难以处理的技术偏见问题同样存在于图书馆与用户的交互过程中，因此，若将交互服务的全部内容交由类 ChatGPT 进行处理，在智慧系统缺乏正确的人类伦理标注与引导、运作仅依据其固有程序规则的情况下，将可能出现向用户传达存在价值观冲突或带有偏见性的内容，产生不必要的争执，造成图书馆与用户之间的紧张关系。

4.3 数据失控：信息安全保护遭受冲击

应用类 ChatGPT 的智慧图书馆将在运营中持续采集用户信息行为数据和新增资源数据，实现语料库数据的不断增长，而对信息的过度采集和不当使用将加剧信息泄露的风险，产生一系列信息安全问题。在用户个人隐私泄露风险上，智慧图书馆需要大量采集用户信息以实现个性化服务，

且所获得的相关数据越多实现个性化服务的质量越高，但此过程极易出现数据过度采集、数据滥用以及数据泄露等侵犯用户个人隐私权的行为。例如，2023 年 3 月 25 日，OpenAI 公司证实了部分网友爆料的 ChatGPT 用户个人隐私与支付信息遭到泄露的情况确有发生[25]。在机构涉密信息泄露风险上，同前所述，若类 ChatGPT 系统对馆藏科研数据或其他涉密数据进行过度采集和使用，在不加以限制数据流动的情况下极可能引发因涉密数据被窃取而导致数据外流。此外，在工作人员与类 ChatGPT 系统的交互过程中，如涉及机构机密，也会被类 ChatGPT 系统采集并使用，在无意中导致涉密数据的泄露。例如：2023 年 4 月 3 日，三星员工错误使用 ChatGPT 将公司涉密信息上传至 ChatGPT 服务器，导致涉密数据泄露海外[26]。

4.4 智能失权：管理决策过度依赖技术

正如前文分析，类 ChatGPT 技术的嵌入能为智慧图书馆的管理决策工作带来极大便利。但随着类 ChatGPT 的智慧决策体系在图书馆工作中的不断深入，图书馆对技术的依赖性不断提高，使得图书馆将不断让渡自身的管理主动性，导致管理权逐渐由人类向人工智能转移，逐渐形成管理主体的智能失权。在这种情况下机器意志将可能取代人的意志，出现智慧图书馆治理中责任主体模糊化的问题。且目前智慧图书馆的技术工作大多委托给第三方科技公司负责，而类 ChatGPT 进行决策的判断与取向是其程序开发者通过标注后的结果，因此这也将导致机器意志的本质仍然是人的主观意志的体现，但这里的人可能并非图书馆工作人员，而是第三方人员。可见，由于图书馆处于技术弱势位置，可能会产生智能失权、黑箱操控等情况，使得资源建设、发展规划等重要决策行为存在受第三方人员影响的隐患。

4.5 人才匮乏：复合型馆员有严重缺口

馆员是智慧图书馆各项服务的基本构成要素之一，同时也是保障智慧图书馆可持续发展的首要因素[27]。随着类 ChatGPT 在智慧图书馆中的应用，智能服务将逐步取代多数图书馆的传统职能，这将使图书馆对馆员职业素养提出更高的要求。一来要求馆员具备丰富的学科知识储备以应对专业化学科服务，二来需具备一定的计算机科学素养，以适应类 ChatGPT 所带来的智慧化环境，因而成为多学科交叉融合背景下的复合型人才，将是生成式人工智能时代智慧图书馆对馆员的职业要求。审视当前智慧图书馆馆员队伍存在严重的人才匮乏现象，尤其缺少具备计算机技术的复合型馆员，这也导致在建设智慧图书馆的过程中不得不依赖于第三方科技公司的技术服务，同时，对技术外包服务的过度依赖也会导致图书馆降低对技术的关注度，使技术应用型人才的招聘与培养遭受轻视，并形成恶性循环[27]。此情况也会为未来智慧图书馆发展埋下人才匮乏的风险。

5 类 ChatGPT 嵌入智慧图书馆的风险防控策略

正确看待类 ChatGPT 为智慧图书馆带来的变革，将类 ChatGPT 的技术优势充分融入智慧图书馆日常工作的同时，综合研判其存在的潜在风险，并制定相应的风险防控策略，能够有效防范类 ChatGPT 嵌入智慧图书馆后可能遇到的各类风险，从而推动智慧图书馆可持续发展。

5.1 构建版权协同保护体系

由于类 ChatGPT 嵌入智慧图书馆后提高了图书馆服务中的知识产权风险，因此，智慧图书馆管理者应当结合类 ChatGPT 的主要技术特征，构建一个具备“技术—管理—法规”三维结构的版权协同保护体系。

（1）引入管控技术。智慧图书馆一方面需要引入数据标注技术，对馆藏资源进行分级标注，将其划分成不同安全等级的数据以便于分层管理。另一方面可以考虑在现有系统中进一步嵌入内容数字版权管理技术（Digital Rights Management），通过数字文件加密、添加数字水印或两者结合的方式，对类 ChatGPT 作用下的智慧图书馆生成内容进行管理与分发[28]。此外，借鉴海外数字内容平台所采用的内容识别技术（Content ID）和版权监视技术（Copyright Watch），开发嵌入类 ChatGPT 的智慧图书馆的版权识别与监视系统，

实现对用户服务过程中侵权行为的识别、制止、提示、记录等自动化处理。

（2）强化版权管理。由于智慧图书馆通常采用第三方科技公司提供的技术支持，因此有必要进一步完善合同条款对其加以制约。例如，对存在潜在版权风险的交互行为进行全程监督、记录及制止，通过对用户 IP、历史行为等数据的分析限制风险用户的权限等。智慧图书馆应当增设知识产权监督委员会，成员包括馆内相关负责人、信息技术专家、知识产权专家等。委员会应履行对智慧图书馆知识产权风险监督与评价的职责，并提供更加完善的知识产权改造方案，以此降低智慧图书馆的知识产权风险。此外，智慧图书馆还应当加强对馆员的知识产权培训，提升馆员的知识产权素养和知识产权保护意识，使馆员能够自主应对日新月异的人工智能环境[29]。

（3）完善相关法规与行业标准。智慧图书馆一方面可以参考现行的知识产权相关法律，制定符合我国法律的智慧图书馆知识产权保护相关行为准则，以此规范用户行为。另一方面，组建智慧图书馆知识产权联盟，针对类 ChatGPT 技术影响下的智慧图书馆知识产权风险制定统一的行业标准及知识产权保护准则。此外，图书馆界还应积极联系立法部门，构建科学的针对性法律法规，将类 ChatGPT 人工智能技术带来的新型知识产权风险纳入司法管理。

5.2 全面构建技术治理机制

类 ChatGPT 所存在的算法黑箱问题，导致人们不仅对其输出结果难以预测，而且无法对其行为边界进行有效控制。因此，对于智慧图书馆而言，应当通过全面构建技术治理机制，以降低风险概率、减轻风险影响。

（1）制定监督与补救机制。智慧图书馆可以针对类 ChatGPT 技术制定一系列监督与补救机制以降低其潜在的技术偏见。在监督方面，智慧图书馆的管理者在深度参与类 ChatGPT 的应用开发过程中，基于行业伦理制度划定类 ChatGPT 的行为边界，使开发过程由“技术方案驱动”向“问题驱动”转变。在补救方面，智慧图书馆的管理者应当围绕类 ChatGPT 嵌入智慧图书馆中潜在风险场景制定较为完善的应急补救预案。当出现错误时，则依据预案及时进行修改、更正、补偿等措施。此外，在技术方面，智慧图书馆的管理者可以考虑在嵌入 ChatGPT 时进一步加入自动升级矫正系统，一旦发生技术偏见问题，通过搜寻并抽取事件全过程的数据，对其进行记录、分析并反馈，从而实现系统的自我学习和升级。

（2）逐步实现“去第三方化”。正如前文分析，当今绝大部分智慧图书馆的技术支撑都需要依赖第三方科技公司，其主要问题在于企业在社会化经营下的逐利性，导致所开发的应用程序意志无法与图书馆意志相统一，在智慧图书馆应用此类程序时易出现技术偏见、决策错误等问题。智慧图书馆的管理者应当积极与所属地方政府、所在机构进行沟通交流，争取获得国有技术团队支持，使图书馆服务场景下的类 ChatGPT 可以从公益角度履行职能。此类做法也可以将图书馆从技术的被动接受者向技术的主导者转换，实现智慧图书馆的“去第三方化”。

5.3 完善数据安全保障体系

从技术原理上看，类 ChatGPT 所采用的“黑箱模型”会导致嵌入类 ChatGPT 的智慧图书馆在运行中隐蔽且频繁地挑战隐私数据、保密数据的安全底线。结合我国《中华人民共和国数据安全法》（以下简称《数据安全法》）第十九条规定来看，智慧图书馆有责任与义务通过确定重要数据保护目录对其进行重点保护，以保障馆藏数据的安全[30]。因此，智慧图书馆应当从制度标准与防范机制的双重视角出发，构建一个适应人工智能环境且更加完善的立体化数据安全保障体系。

（1）制定数据合规标准。智慧图书馆存在数据安全问题的实质是其内部防范与应对数据安全风险的能力不足[31]。因此，根据类 ChatGPT 的应用特点和技术特征进一步优化制度数据合规标准、强化数据全流程的合规与监管，是完善智慧图书馆数据安全保障体系的重要基础。例如，将《中华人民共和国公共图书馆法》《中华人民共和国网络安全法》《数据安全法》等国内法律法规，

《人工智能法案》《人工智能权利法案蓝图》等国外法律以及《数字图书馆安全管理指南》《数字图书馆资源管理指南》等行业规范文件相结合，构建最符合我国实际情况的智慧图书馆数据合规标准，以此全面提升智慧图书馆数据安全保障能力。

（2）建立风险防范机制。尽管类 ChatGPT 的深度运用实现了对智慧图书馆服务的全类型、全过程以及全时段的覆盖，但是同样也导致数据安全风险存在于使用的全过程当中。因此，智慧图书馆的管理者应当建立一系列具体的风险防范机制，用以强化智慧图书馆的数据安全保障体系。一方面，可以开发自动化数据安全风险预警系统并将其嵌入智慧图书馆，以实现数据安全监管常态化、智能化，借助系统的自我监督搭建起智慧图书馆监测与预警机制。另一方面，可以针对数据生命周期的每个环节，通过数据脱敏、数据备份、数据加密以及用户访问权限限制等方式深化智慧图书馆数据管理，以此降低数据安全风险。

5.4 创新组织领导管理机制

类 ChatGPT 嵌入智慧图书馆是数智时代下图书馆智慧化建设的必然选择，同时在智慧图书馆利用类 ChatGPT 实现全面升级的过程中，技术的更新也将反向作用于图书馆的管理模式，对其管理机制提出新的要求。因此，为了避免管理决策过度依赖技术，智慧图书馆的管理者应当创新领导管理机制，使其与智慧化环境相匹配，以保障智慧图书馆在“数智时代”的可持续发展。

（1）创新领导选拔机制。智慧图书馆的领导班子是否具备足够的算法素养和智慧管理理念，是保障智慧图书馆嵌入类 ChatGPT 后能否持续发展的关键。虽然领导选拔权利不属于图书馆，但图书馆领域应结合实际，从专业角度制定基于人工智能环境特征的领导职位专业门槛，创新干部选拔任用机制，与政府积极沟通争取采纳，从而避免出现外行管理内行的现象。一方面，引入首席信息官（Chief Information Officer，CIO）和首席数据官（Chief Data Officer，CDO）制度，为信息技术、数据管理、保密管理等专业人才设置专属高级管理岗位，由其全面负责智慧图书馆的信息系统构建与数据安全保障工作。另一方面，在智慧图书馆领导干部的考核选拔过程中，应当在传统选拔机制基础上增加数据治理能力要求，并给予一定权重。通过考核倒逼智慧图书馆的领导团队进一步提升人工智能环境下的管理视野和技术能力，以适应数智时代下的智慧图书馆管理工作。

（2）调整馆内组织结构。结合类 ChatGPT 嵌入智慧图书馆的应用情况来看，调整馆内组织结构，制定机构协同治理体系，对于减少“智能失权”风险具有重要作用。一方面，智慧图书馆的管理者应当加速部门扁平化设置，通过分析嵌入类 ChatGPT 智慧图书馆中馆员的主要职能与活动，重新调整部门。同时，还需依据多元需求设立不同的专项小组，且组长仅对接管理层，进一步实现扁平化管理下集权与分权相结合的组织结构。另一方面，智慧图书馆的管理者应当尽可能推动机构、部门、专项小组三者之间建立多元协同关系，进一步强化所有主体的责任义务，充分发挥各自的职能作用，共同构建理想的智慧图书馆组织生态环境。

5.5 推动技术人力资源建设

馆员的职业技能与前沿技术的匹配性对于智慧图书馆的可持续发展至关重要。随着类 ChatGPT 在智慧图书馆中的深度应用，馆员角色同样随之发生转变。因此，智慧图书馆应当通过推动技术人力资源建设，组建符合数智时代环境需求的专业馆员队伍。

（1）制定馆员职业技能标准。在应用类 ChatGPT 实现高度智能化的智慧图书馆环境下，图书馆员需要具备与智慧环境相匹配的知识与技能[32]。因此，各级文化管理部门、图书馆学（协）会、图书情报工作委员会以及各智慧图书馆的管理者可以考虑制定一套人工智能背景下的图书馆员职业技能考核标准，在业务技能、服务技能、管理技能等传统指标上加入数据管理应用技能、数据挖掘分析技能、用户需求分析技能、程

序开发优化技能等新指标，设立相关的技能水平考试并颁发证书，以此提升馆员的学习积极性和专业性。

（2）构建全方位的培养机制。为应对类 ChatGPT 带来的职业挑战，智慧图书馆可以参考前述的馆员职业技能标准，建立全方位的内部培养体系。一方面，智慧图书馆可以邀请人工智能、图书情报等领域的专家、学者开设讲座或主题报告，其内容包括人工智能概述、馆藏资源建设、智慧图书馆系统实操、业务经验交流等，以此提升馆员在类 ChatGPT 技术上的理论与实践能力。另一方面，图书馆应为馆员提供更多“走出去”的交流学习机会，采用异地业务交流、挂职锻炼、访问学者等方式进入其他图书馆和人工智能企业，从而了解当前的技术走向、应用现状、发展趋势以及学习工作所需的技能。

（3）增加岗位吸引力，引进高层次人才。长期以来图书馆难以招聘到高层次技术型人才，其原因在于人才选拔方式、人才利益追求、人才额外收益等方面存在问题[33]。因此，智慧图书馆应通过改革传统用人机制与人才待遇体系吸引国内外高层次人才加入。一方面，可以考虑放宽硬性指标，降低在学历、年龄等方面的要求，但要从应聘者的技术能力、项目背景、职业认知等方面考察其是否能真正胜任人工智能环境下智慧图书馆的工作要求。另一方面，图书馆可以通过创新绩效二次分配办法，以提高人才岗位实际待遇。此外，应当尽可能制定一些符合人才需求的激励政策，如晋升路径、带薪休假、创业机会等，以此为高层次人才提供全方位的保障。

[参考文献]

[1] 郭全中，张金熠．ChatGPT 的技术特征与应用前景［J］．中国传媒科技，2023（1）：159－160.
[2] 冯志伟，张灯柯，饶高琦．从图灵测试到 ChatGPT——人机对话的里程碑及启示［J］．语言战略研究，2023，8（2）：20－24.
[3] 尹克寒．ChatGPT 的发展对情报信息工作的影响及启示［J］．图书馆理论与实践，2023（3）：15－22.
[4] Kohinoor M D，Emily S．Cross．The Computer a Chore-ograp her? Aesthetic Responses to Computer-generated Dance Choreography［EB/OL］．［2023－05－18］．https://psyarxiv.com/yvgxk/.
[5] 储节旺，杜秀秀，李佳轩．人工智能生成内容对智慧图书馆服务的冲击及应用展望［J］．情报理论与实践，2023，46（5）：6－13.
[6] 李书宁，刘一鸣．ChatGPT 类智能对话工具兴起对图书馆行业的机遇与挑战［J］．图书馆论坛，2023，43（5）：104－110.
[7] 吴若航，茆意宏．ChatGPT 热潮下的图书馆服务：理念、机遇与破局［J］．图书与情报，2023，210（2）：34－41.
[8] 蔡子凡，蔚海燕．人工智能生成内容（AIGC）的演进历程及其图书馆智慧服务应用场景［J］.图书馆杂志，2023，42（4）：34－43，135－136.
[9] 郭亚军，郭一若，李帅，等．ChatGPT 赋能图书馆智慧服务：特征、场景与路径［J］．图书馆建设，2023（2）：30－39.
[10] 李立睿，张嘉程，魏银珍，等．智能机器人赋能图书馆服务：内涵、特征与实施路向［J］．图书馆学研究，2022，526（11）：10－18.
[11] Vaswani A，Shazeer N，Parmar N，et al．Attention is All You Need［J］．Advances in Neural Information Processing Systems，2017（30）：1－20.
[12] 张重毅，牛欣悦，孙君艳，等．ChatGPT 探析：AI 大型语言模型下学术出版的机遇与挑战［J］．中国科技期刊研究，2023，34（4）：446－453.
[13] 周懿琼，蔚雷．副省级公共图书馆数字资源建设情况调研分析［J］．数字图书馆论坛，2022（11）：60－66.
[14] 王建磊，曹卉萌．ChatGPT 的传播特质、逻辑、范式［J］．深圳大学学报（人文社会科学版），2023，40（2）：144－152.
[15] OpenAI．GPT-4［EB/OL］．［2023－03－21］．https://openai.com/research/gpt-4.
[16] 丁明春，任恒，叶路扬．论智慧馆员职业能力的核心要素及其提升策略［J］．图书馆理论与实践，2022（2）：33－39.
[17] 许春漫，陈廉芳．高校图书馆智慧服务模式下智慧馆员队伍的建设［J］．情报资料工作，2014（1）：87－91.
[18] 金敏婕．融入、提升、超越——智慧图书馆员

素养与价值［J］. 图书与情报，2014（6）：130－133.
［19］司莉，马小景. 元宇宙视角下虚拟数字人赋能图书馆用户服务研究［J/OL］. 图书馆建设：1－8.［2023－04－17］. http://kns.cnki.net/kcms/detail/23.1331.g2.20221221.1317.003.html
［20］陈昊琳，刘亭亭. 公共图书馆业务管理变革趋势解读——《国家图书馆业务管理机制》读后［J］. 图书馆工作与研究，2019，279（5）：31－36.
［21］徐毅，钱智勇. 面向国家文化数字化战略的图书馆创新实践——以张謇学濒危稀见文献数字化保护和利用为例［J］. 新世纪图书馆，2022，315（11）：20－29.
［22］李玉海，金喆，李佳会，等. 我国智慧图书馆建设面临的五大问题［J］. 中国图书馆学报，2020，46（2）：17－26.
［23］丛立先. 人工智能生成内容的可版权性与版权归属［J］. 中国出版，2019（1）：11－14.
［24］熊琦. 人工智能生成内容的著作权认定［J］. 知识产权，2017（3）：3－8.
［25］裘雯涵. 涉嫌数据泄露，意大利暂停 ChatGPT［N］. 解放日报，2023－04－05（6）.
［26］科创板日报. 三星员工被曝不当使用 ChatGPT 半导体机密数据直传美国［EB/OL］.［2023－05－01］. https://baijiahao.baidu.com/s?id=1762148832034469161&wfr=spider&for=pc.
［27］袁豪杰. 高校图书馆人力资源管理［J］. 现代情报，2008（8）：107－109.
［28］Becker E，Buhse W，Günnewig D，et al. Digital Rights Management：Technological，Economic，Legal and Political Aspects［M］. Berlin：Springer Verlag，2004：101－112.
［29］周静，张立彬，谷文浩. 我国高校图书馆知识产权信息服务的现状与思考［J］. 图书情报工作，2019，63（21）：35－46.
［30］邓灵斌.《数据安全法（草案）》解读及我国图书情报界的对策建议［J］. 情报杂志，2020，39（12）：83－87.
［31］盛豪杰. 图书馆数据合规：缘起、原理及具体构建［J］. 图书馆，2022，334（7）：1－9.
［32］邓李君，杨文建. 国外图书馆员职业能力标准的特征解析及启示［J］. 国家图书馆学刊，2023，32（1）：3－15.
［33］郝晓玫. 浅析图书馆人才危机的原因及对策［J］. 图书馆理论与实践，2012（9）：11－13.

［作者简介］刘凌宇（1990— ），男，丽水学院图书与信息服务中心馆员，研究方向：智慧图书馆与数据智能；徐中阳（1994— ），男，杭州电子科技大学管理学院博士研究生，研究方向：信息计量与科教评价，智慧图书馆与数据智能。
［收稿时间］2023－09－18　［责任编辑］张静婕

（上接第 44 页）

［20］阮可. 现代公共文化服务体系：理论与浙江实践［M］. 杭州：浙江大学出版社，2014：126－127.
［21］象山县人民政府. 象山溪里方村文化礼堂获评“最美文化空间”［EB/OL］.［2023－11－06］. http://www.xiangshan.gov.cn/art/2023/3/27/art_1229676067_59147512.html.
［22］许欢. 我国公共阅读空间的建立与现代图书馆发展研究［J］. 图书馆建设，2010（1）：109－112.
［23］李建霞，文卫华. 我国公共阅读空间的兴起与发展趋势探析［J］. 出版广角，2019（8）：19－21.
［24］戴艳清，孙英姿. 城市新型公共阅读空间可达性研究——基于长沙市中心城区新型公共阅读空间的调查［J］. 图书馆学研究，2023（1）：2－11.
［25］黄佩芳. 我国城市公共阅读空间建设特点与模式选择［J］. 图书馆，2019（3）：90－94.
［26］中华人民共和国国家发展和改革委员会.“十四五”规划《纲要》名词解释之三丨高质量发展［EB/OL］.［2023－11－17］. https://www.ndrc.gov.cn/fggz/fzzlgh/gjfzgh/202112/t20211224_1309252.html?state=123&state=123&state=123.

［作者简介］甘佩玄（1998— ），女，南京大学信息管理学院 2022 级硕士研究生，研究方向：公共文化服务；陈雅（1965— ），女，博士，南京大学信息管理学院教授，博士研究生导师，研究方向：公共文化服务。
［收稿日期］2023－11－30　［责任编辑］闫东芳