AIGC 在图书馆服务创新中的安全风险与治理策略研究

马晓亭

摘　要　随着人工智能生成内容（AIGC）技术的发展，图书馆可以利用AIGC技术实现在智能化管理、读者服务模式和服务路径等方面的新变革，但图书馆AIGC应用服务模式的多样性、大数据环境的复杂性和个性化定制的唯一性，也给图书馆AIGC技术的落地与应用实践带来了严峻的安全风险问题。论文首先研究了AIGC技术在图书馆智慧服务中的应用场景，探讨了其对图书馆服务模式变革和服务路径创新的智慧赋能方式。然后，研究了图书馆AIGC应用实践面临的安全风险与威胁，以及这些风险与威胁对图书馆智慧服务模式变革和AIGC智慧阅读场景、内容生成的影响。最后，构建了图书馆AIGC风险防控系统结构与检测流程，并针对图书馆面临的安全风险与威胁提出了相应的防范策略。该策略结合图书馆AIGC阅读服务中数据生命周期管理流程，应用于生成系统数据输入、AIGC模型合成、生成的场景与内容、读者AIGC阅读流程等，可以为读者提供安全、高效、个性化和可扩展的智慧阅读服务。

关键词　人工智能生成内容；图书馆服务创新；安全风险；治理策略
分类号　G250.76
DOI　10.16810/j.cnki.1672-514X.2024.12.005

Research on Security Risks and Governance Strategies of Artificial Intelligence Generated Content (AIGC) in Library Service Innovation

Ma Xiaoting

Abstract　With the development of Artificial Intelligence Generated Content(AIGC) technology, libraries can use AIGC technology to achieve new changes in library intelligent management, reader service models, and service paths. However, the diversity of library AIGC application service models, the complexity of big data environments, and the uniqueness of personalized customization also bring huge and serious security and legal issues to the implementation and application practice of library AIGC technology. Firstly, this paper studies the application scenarios of AIGC technology in library intelligent services, and explores the intelligent empowerment methods of AIGC technology for the transformation of library service modes and innovation of service paths. Then, it studies the security risks and threats faced by library AIGC application practices, as well as the impact of these risks and threats on the transformation of library intelligent service modes and the AIGC intelligent reading scenarios and content generation. Finally, it constructs the structure and detection process of the library AIGC risk prevention and control system, and puts forward corresponding prevention strategy for the security risks and threats faced by libraries. This strategy integrates the data life-cycle management process in the AIGC reading service of the library, comprehensively supervises the data input of the generation system, the synthesis of AIGC models, the scenarios and content generated, and the AIGC reading process of readers, etc. The strategy can provide safe, efficient, personalized and scalable AIGC intelligent reading services for readers.

Keywords　Artificial intelligence generated content. Library service innovation. Security risks. Governance strategies.

0　引言

近年来，伴随着人工智能生成内容（Artificial Intelligence Generated Content，以下简称 AIGC）技术的快速发展，AIGC 技术重新塑造了用户数字信息的生产、服务和消费方式。用户可以利用它生成对抗网络（GAN）和大型预训练模型等人工智能技术，结合已有数据寻找规律，并通过适当的泛化能力，生成相关场景与内容[1]。AIGC 技术为用户生成的内容不仅仅是常见的文本、音频、图像和视频等，而且也包括策略、剧情、

训练数据等内在逻辑内容。图书馆的用户服务具有读者需求多样、个性化定制、适时动态推荐和阅读收益率要求高的特点，因此，AIGC 技术已悄然引领着一场深刻的服务变革，将重塑图书馆读者个性化服务的模式和路径 [2]。

AIGC 是由人工智能、云计算、大数据和 5G 等技术融合而来，AIGC 在增强图书馆服务模式科学性、推动服务内容生成效率、提升读者阅读收益和用户满意度的同时，也带来了新的应用安全和法律风险，面临着技术复杂度高、训练数据海量和多模式、服务应用路径多样性、监管有限等问题，导致图书馆在 AIGC 应用中面临着诸如大数据安全保护、读者隐私保护、数据防伪造、AIGC 技术生成内容的诚信风险、AIGC 应用版权保护、算法偏见和歧视等风险。因此，在牢牢抓住 AIGC 为图书馆带来服务模式变革和服务路径重塑机遇的同时，如何有效防范、化解 AIGC 技术带来的安全和法律风险，是图书馆从根本上杜绝 AIGC 应用风险，提升图书馆智慧服务体系变革和新服务路径整体安全性面临的一个重大课题。

1　图书馆 AIGC 的应用场景、面临的风险与威胁

1.1　图书馆 AIGC 的应用场景

AIGC 作为一种新型的生产方式，可利用人工智能技术动态生成图像、文本、语音、视频和虚拟的现实场景，其自动生成内容具有精确、高效、快速、低成本、多样化和跨越多种模式形态的特点，图书馆可利用 AIGC 技术实现图书馆智能化管理、读者服务模式变革和服务路径创新。在图书馆智慧化管理中，图书馆利用 AIGC 技术实现馆藏资源和管理资源的虚拟 3D 呈现，大幅提升图书馆馆员组织、管理馆藏资源和管理读者服务系统的效率。在图书馆智慧服务模式变革中，图书馆可利用 AIGC 技术实现服务系统知识组织、读者智慧咨询、馆员信息素养教育、读者个性化虚拟阅读定制和读者的 AI 智慧创作等，实现了服务模式从智能化向智慧化的转变。从 AIGC 技术在图书馆实践应用的生成模式来看，主要涉及文本生成、音频生成、图像生成、视频生成和跨模态生成等 5 大类，具体的应用场景和作用方式如表 1 所示 [3]。

表 1　图书馆 AIGC 应用模式分类及实践路径实例表

模式分类	具体分类	应用路径与具体应用实例
文本生成内容	文本理解	文本内容、语义、情感和隐含思想的分析与解释
	结构化写作	特定情景下的个人总结报告、新闻撰写等
	非结构化写作	创意性写作、营销文本策划等
	交互性文本	机器人客服、聊天问答、读者个性化场景学习
音频生成内容	语音仿真	云听、有声读物制作、语音播报和导航等
	语音机器人	机器人的智慧客服、翻译、老师与主播等
	音频创作	作曲、播客、电影、游戏、人声录制和整体混音等
图像生成内容	图像编辑与创作	图像编辑、分辨率提升、动画与教学视频制作等
	2D 图像转 3D	游戏开发、教育、科研、医学影像、虚拟呈现等
视频生成内容	图像增强修复	视频插帧、视频细节增强和损坏帧修复
	转换视频风格	教学视频风格转换、影像成像效果增强等
	面部动态转换	AI 换脸
	视频动画创作	制作动画视频教学片、精彩视频回顾、AI 虚拟人
跨模态生成内容	文本生成图像	文字生成图书插图或图片
	文本生成视频	文字生成教学视频、娱乐电影或短视频、广告等
	图像 / 视频转文字	搜索引擎或问答系统
	文字转代码	自动化生成电子邮件、编写代码

AIGC 技术虽然为图书馆智慧服务模式变革和服务实践路径创新提供了强大的技术支持，使图书馆的服务模式由传统的单一模式向智慧多样化转变，服务路径也由千人一面向读者个性化定制方式转变。但是，AIGC 技术在极大提升图书馆服务高效性、经济性、个性化、实时性和读者阅读收益的同时，也给图书馆 AIGC 技术的应用实践带来了严峻的安全、法律问题，是数智化环境下智慧图书馆转型升级与持续发展必须关注的一个重要问题 [4]。

1.2　图书馆 AIGC 应用实践面临的风险与威胁

图书馆在用户服务模式转型和服务路径高效、多样化探索中，利用 AIGC 技术的机器自主学习与自然语言处理技术，通过对已采集海量数据的总结与语言规则分析，以及预训练大模型、生成式对抗网络等方法，结合读者个性化智慧阅读活动高效、实时、精确、个性化和低成本的需求，智慧、动态、实时和快速地生成读者阅读场景与内容，极大地提升了图书馆智慧服务的科学性、多样性、个性化定制和读者阅读收益率 [5]。

但是，当前 AIGC 技术具有飞速发展和较少实践应用经验的特点，在给图书馆智慧赋能的

同时也带来了诸如数据安全管理、虚假信息、意识形态偏见和侵犯用户著作权等风险。虽然我国在保护国家数据安全、维护公民数据权益和构建安全网络环境中出台了《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》《中华人民共和国计算机信息系统安全保护条例》《中华人民共和国著作权法》等法规，但仍不能完全、彻底地发现与解决当前图书馆在 AIGC 技术应用实践中产生的新问题，图书馆 AIGC 应用实践过程面临着新的安全风险与威胁，直接关系着图书馆服务模式智慧转型能否成功和用户服务路径实践的有效性，是当前图书馆 AIGC 应用面临的一个重要问题 [6]。

1.2.1　图书馆 AIGC 系统框架结构与 AIGC 应用内容的安全问题

由表 1 可知，我们将图书馆 AIGC 实践按照应用模式分类，可划分为文本生成内容、音频生成内容、图像生成内容、视频生成内容、跨模态生成容等 5 个一级应用模式部分，依据具体应用实践路径分类可划分为文本理解、结构化写作、非结构化写作、交互性文本、语音仿真、语音机器人、音频创作、图像编辑与创作、2D 图像转 3D、图像增强修复、转换视频风格、面部动态转换、视频动画创作、文本生成图像、文本生成视频、图像 / 视频转文字、文字转代码等 17 个二级分支应用部分 [7]。在图书馆多样化的 AI 应用路径中，存在的首要风险为是否部署高效、精确的内容生成安全管理机制，能否对中、英文提示词和违规内容进行拦截过滤，并在 AIGC 应用的初始状态图书馆能否通过对创作素材的过滤而防范违规内容的生成。此外，还要防止 AIGC 基于合规素材的二次创作成果而间接生成违规内容，从源头上严格防范“毒教材”等类似事件的发生。

图书馆在利用 AIGC 技术进行读者个性化定制内容的生成中，如何防范不法分子利用深度伪造技术生成虚假的阅读文本、音频和视频，以及通过 AI 换脸、语音模拟、人脸合成、视频生成场景等，实施反动政治宣传、宣扬色情暴力和教唆犯罪等，也是图书馆 AIGC 应用面临的一个严峻问题。

为了保证 AIGC 生成内容精准、实时和逼真，图书馆会尽可能地提供海量、真实和详细的数据进行机器学习与训练，这些数据可能涉及国家、企业和公民的隐私数据与商业机密，高密级训练数据的泄露可能会导致图书馆 AIGC 应用泄密。以 ChatGPT 为例，其使用条款明确规定除非用户要求 OpenAI 不对其输入和输出内容进行使用，否则 OpenAI 拥有对任何用户输入和输出内容的广泛使用权，以达成改善 ChatGPT 的目的。因此，如何采用法规和技术问题防范泄密，是图书馆 AIGC 应用要关注的一个重要问题。

1.2.2　侵害用户著作权与个人隐私安全问题

图书馆 AIGC 技术的实践应用，是一个基于海量数据的采集、存储、管理、机器训练和内容生成的复杂过程，未经用户授权的非法访问、数据泄露或者丢失、数据的过度与恶意使用，都可能造成严重的 AIGC 应用安全问题，可能会侵害用户的著作权益和个人隐私安全 [8]。

当图书馆 AIGC 应用中机器通过较深度的学习与训练后，其生成的场景与内容将更加真实和生动，且 AIGC 应用实践呈现多样和低成本的特点，海量、多模式的生成内容具有大规模、分散、流动和隐蔽的特点，仅靠图书馆一方难以实现实时、动态和精确的监管。

海量训练数据是图书馆 AIGC 服务实践科学、高效的前提和基础，而图书馆海量训练数据主要来源于自有数据库、向第三方采购买和网络爬取，而这种在互联网上利用专门软件随意爬取和使用的数据，可能会侵犯他人数据隐私和所有权益 [9]。

图书馆 AIGC 应用需要依托海量的文本、语音和视频数据，通过对机器高强度的训练、学习而生成优质的内容。一方面 AIGC 应用所依赖的基础海量数据可能包含他人未经许可使用的作品，会导致对他人著作成果的侵权；另一方面即使生成内容中包含了 AIGC 算法工程师的劳动力，但在版权申请中也会因为作品与他人作品存在“实质性相似”，或者不具备独创性而无法获得著作权。

1.2.3　AIGC 模型缺乏透明度与人工智能算法存在歧视风险

图书馆在设计 AIGC 模型时可能会过度关注模型的功能和属性，而 AIGC 模型功能的实现仅需要调用其功能和属性即可，而不去关心这些功能与属性是如何实现的，导致部分 AIGC

模型缺乏透明度、可解释性和可控性。图书馆在应用这些存在“黑盒子”特征的 AIGC 模型时，系统分析师难以预测和信任这些 AIGC 模型实际生成的场景与内容 [10]。

图书馆 AIGC 模型在用户服务模式重构和服务路径实践时存在一定的不可控性，同样的训练数据输入不同的 AIGC 模型系统，AIGC 系统输出数据所生成的阅读场景和内容，会对具有不同文化背景、价值观点、民族特征、生活习惯、宗教信仰和阅读习惯的读者产生巨大差异，因此，图书馆对基于特定输入数据的 AIGC 生成场景与内容的人工审查必不可少。此外，存在偏见的训练基础数据可能会导致 AIGC 模型生成偏见的读者阅读场景与内容。

图书馆 AIGC 模型构建的科学性和训练数据真实性是关系 AIGC 系统生成内容有效的关键。当前用户在使用 ChatGPT 时，普遍发现它更擅长解答文科类的问题，而对基础知识和逻辑性要求较高的理工科类问题问答准确率较低，甚至在某些特殊的应用场景下会产生奇怪、不准确、缺乏创新、侵犯隐私与知识产权、错误或违背伦理道德的答案。另外，图书馆 AIGC 系统结构的不完善性可能会导致生成的读者阅读场景、文本、图像和音频数据与原始数据存在巨大差异，误导读者获取不相关、不精确甚至错误的知识。

1.2.4　AIGC 训练数据污染与算法滥用风险

基于 AIGC 的图书馆读者智慧阅读模式与用户服务路径具有复杂、多样性的特点，大量 AIGC 机器人依据读者智慧阅读需求不间断地创造、合成、转发和训练数据，可能会在图书馆内部系统形成 AIGC 机器人互动网络。这些在图书馆 AIGC 机器人中传输的数据存在着大量的错误、噪声与不相关数据，并在 AIGC 推荐算法的分发逻辑下进一步分发、扩散，会造成图书馆 AIGC 应用环境的信息污染，严重影响 AIGC 生成内容的准确性和效率 [11]。

此外，如果图书馆 AIGC 机器人训练数据被严重污染或网络被黑客入侵，在错误训练数据的影响下 AIGC 机器人可能会生成相似的错误训练内容而误导读者。特别是在关系到重大国际事件、突发事件、历史真相、国家决策和价值观的重大生成内容中，会因为“烟雾遮蔽”（AIGC 机器生产的海量虚假、伪造内容数据遮掩人工生产的有限真实内容数据，进而干扰读者的正确视听）或“标签劫持”（通过发布大量与标签主题无关内容，来“干扰”或“占有”原有标签，以及用负面评论战略性地淹没原本主题，降低真实标签舆论影响的一种网络攻击行为）而影响读者智慧阅读信息接收的真实性、准确性与效率，甚至给读者传输错误、反动和违法的信息 [12]。

算法滥用是另一个威胁图书馆 AIGC 应用的严重问题。一些黑客可能会利用 AIGC 模型强大的机器学习能力，生成计算机病毒攻击 AIGC 机器人；或者采用非法手段直接攻击、修改和窃取 AIGC 算法，使 AIGC 按照黑客的意图生成错误的内容，并利用错误数据欺骗人工智能系统使其做出错误的判断。此外，如何确保 AIGC 训练数据公正、公平和中立，防止图书馆 AIGC 生成的阅读内容产生偏见或歧视性结果，也是影响读者智慧阅读满意度和收益率的一个重要因素。

1.2.5　AIGC 数据应用的合法性与合规性风险

海量、正确和强相关性数据是确保图书馆 AIGC 训练科学、智慧和逼真的前提。AIGC 系统训练数据不仅包括图书馆的自有服务应用数据和外购数据，还应包括其他政府组织、企业团体、图书馆行业、社交媒体、公共论坛和私人拥有的数据。例如，图书馆在采集数据时，如果部分网站严禁利用爬虫技术爬取数据或对数据的使用范围有相关应用限制，或者个人要求图书馆删除已应用于 AIGC 机器训练的私人数据，则可能会影响图书馆 AIGC 应用有效性。此外，如何确保图书馆采集的相关数据不被非法用于大数据深度分析国家机关、企事业单位的相关机密或公民隐私数据，也是图书馆 AIGC 应用所面临的一个难题。

为了增加图书馆 AIGC 应用的智慧水平，AIGC 机器训练数据应包括多个行业领域和全球不同地域国家数据，图书馆在使用境外数据进行 AIGC 机器训练时，不同国家和地域有自己相关的数据保护法律和使用要求。欧盟规定个人数据可以在欧盟成员国之间自由流动，没有任何限制，但是非欧盟成员国想要从欧盟成员国获取个人信息数据，就需要达到适用性、充分性标准。全球不同国家较大差异性的数据管理法规，也给我国图书馆界跨境数据采集、使用造成了极大的不便与风险 [13]。

2　图书馆 AIGC 应用安全威胁与风险防范策略

2.1　加强我国 AIGC 应用立法、监管和图书馆 AIGC 行业应用规范的制定

伴随 AIGC 技术的快速发展和应用领域的不断拓展与深入，全世界 AIGC 应用风险事件频发，各国开始了对人工智能领域立法、监管的探索与实践。2020 年 11 月美国发布了《人工智能应用监管指南》，该指南指导美国政府如何制定人工智能监管政策，其重点为确保监管规则不阻碍人工智能的发展。2022 年 10 月，白宫科技政策办公室发布《人工智能权利法案蓝图》，核心内容为五项基本原则，其中将公平和隐私保护放在首要位置。2023 年 1 月，NIST（美国国家标准与技术研究院）正式发布《人工智能风险管理框架》，该框架可由相关机构自愿选择使用，旨在提供设计、开发、部署和使用人工智能系统的指南，增强人工智能可信度，降低人工智能技术应用的风险。欧盟在人工智能实践过程中在数据和隐私保护上最为保守，2023 年 12 月 8 日，欧盟就《人工智能法案》达成协议，该项法案旨在通过全面监管人工智能，为这一技术的开发和使用提供更好的条件。2024 年 2 月 2 日，欧盟 27 国代表一致支持《人工智能法案》文本。3 月 13 日，欧洲议会通过了《人工智能法案》，《人工智能法案》是全球首个 AI 监管法案。2023 年 3 月 29 日，英国政府发布了人工智能新监管框架的提案——《一种支持创新的人工智能监管方法》（白皮书），它的目标是“提供一个清晰的、有利于创新的监管环境”。英国针对人工智能领域的监管框架将基于五个关键原则：一是安全、保障和稳健性；二是适当的透明度和“可解释性”；三是公平；四是问责制和治理；五是可竞争性和补救。国际标准化组织（ISO）也已成立人工智能的分委员会 SC42，它负责人工智能标准化工作，重点围绕数据质量与治理、可信与安全展开研究工作。

为了鼓励和保护我国互联网产业与新兴科技的发展，防止因超前立法而阻碍科技的发展和创新，我国在诸如互联网产业及“互联网 +”这类新技术新事物、新业态的发展初期，就明确提出“最大限度减少事前准入限制，加强事中事后监管”[14]。依据《中华人民共和国网络安全法》《中华人民共和国数据安全法》《个人信息出境标准合同办法》等，结合 AIGC 类算法监管，国家网信办与工信部、公安部于 2022 年 11 月 25 日发布了《互联网信息服务深度合成管理规定》，就深度合成技术作出了针对性合规指引。2023 年 8 月 15 日施行的《生成式人工智能服务管理暂行办法》，从生成式人工智能服务提供者的算法设计与备案、训练数据、模型，到用户隐私、商业秘密的保护，监督检查和法律责任等方面提出了相关要求，同时中明确了对生成式 AI 产业的支持和鼓励态度[15]。

伴随着 AIGC 技术的不断发展与产业升级，AIGC技术将会构建崭新的图书馆用户服务模式，依据读者个性化阅读定制需求，创新生成更多的读者阅读场景和应用产品，并且这种创新是基于全球不同国家、地域和民族数据的共享与交流。因此，在我国密切关注世界各国 AIGC 立法趋势并制定出符合我国国情的监管法律同时，图书馆界也应该结合我国法律新规、新动向和 AIGC 应用实际，制定出相应的行业规范并及时动态调整，才能保证图书馆 AIGC 应用安全、高效、经济和可控。

2.2　图书馆应加强对 AIGC 产品服务商资质和应用产品安全性的审核

图书馆 AIGC 的应用与实践主体，主要由图书馆、读者、AIGC 深度合成服务系统提供商、AIGC 应用数据提供者等 4 个个体组成，因此，在 AIGC 应用中应坚持“责任共担模型”的原则，从数据的采集、分析、处理、内容生成、用户推荐和服务应用流程入手，由图书馆和专业 AIGC 安全管理机构共同评估、发现、整改和防范相关实践风险。

对于图书馆而言，应在 AIGC 产品上架之前联合具备 AIGC 安全资质的公司，全面检查拟上架 AIGC 应用产品的合法、合规性。这个检查应覆盖 AIGC 产品实践的全生命周期流程，包括机器训练数据的选择、算法设计及审核、内容合规、知识产权保护、生成内容标识、算法备案、安全评估等，特别要重点检查 AIGC 系统的安全性、可解释性、生成场景的合法性等方面[16]。对于未获取 AIGC 研发许可证和相应《计算机软件著作权证书》的应用系统，图书馆应按照不合规产品下架处理。此外，图书

馆还应吸收和培养 AIGC 数据管理与系统应用的专业人才，协助图书馆取得相应的安全认证资格，开展 AIGC 应用实践。

AIGC 深度合成服务系统提供商负责图书馆 AIGC 应用产品的设计、开发和运营全过程，是 AIGC 产品安全性保证的直接责任人，负责对 AIGC 产品全生命周期流程的安全管理与防护。AIGC 深度合成服务系统提供商应结合我国《生成式人工智能服务管理暂行办法》要求，在全国互联网安全服务管理平台进行资质审查和备案，获得相应的 AIGC 研发和应用 ICP 许可等资质。在开发相应的图书馆 AIGC 应用系统时，应对系统进行科技伦理审查和法律风险评估，并申请相关的《计算机软件著作权证书》，防止软件侵权事件的发生。在 AIGC 系统的训练中，应严格审核机器训练数据的数据质量、可用性和相关性，并将生成内容算法备案，确保算法模型科学和合理使用 [17]。

读者是既是图书馆 AIGC 成果的服务对象，也是 AIGC 系统生成内容输入数据的选取与供给者，读者在 AIGC 系统输入的文字、语音、图像和视频等原始数据的合法、合规性，将直接关系生成内容的合法、可用性。因此，图书馆应加强与读者 AIGC 阅读活动相关的用户注册、实名认证、生成内容安全合规性检查、端口输入数据安全、个人隐私保护、生成内容安全性综合评估的管理，防止读者非法使用 AIGC 生成阅读内容。

2.3 加强图书馆 AIGC 应用相关版权保护

图书馆 AIGC 应用是建立在多模态之上的人工智能技术，基于机器的深度自主学习技术，能够学会语言的语法、语义并理解上下文，从而更好地生成与训练数据相似的内容 [18]。图书馆 AIGC 系统可以同时理解语言、图像、视频、音频等，能够完成单模态模型无法完成的任务，比如给视频添加文字描述、结合语义语境生成新的图片、语音、视频等，可根据读者阅读需求为读者提供崭新的个性化阅读场景、内容和服务路径。

数据是图书馆 AIGC 系统训练和内容生成的基础，其 AI 模型的数据来源主要有公共共享数据集、互联网网站、自有数据、众包数据和合成数据等几大类，除了自有数据和合成数据外，其他数据来源都可能有版权的争议。如果图书馆仅仅通过数据去噪与解码处理，或者 AIGC 生成的内容与数据原主体有“实质性相似”，则可能会侵犯数据或者作品原始拥有者的版权。我国著作权法对作品是否侵权，主要从 4 个方面进行判断，即：作品在文学、艺术和科学领域内；具有独创性；能以一定形式表现；属于智力成果，并且所有权保护的主体是中国公民、法人或者非法人组织 [19]。因此，对于明确拥有所有权的数据图书馆应通过购买、授权的方式获得数据的使用权，而对于 AIGC 智能创作的内容，则应以读者在阅读活动中 AIGC 系统生成的内容是否体现了 AIGC 系统开发者或系统使用者的创新性为依据，判定版权所有对象。

《中华人民共和国著作权法》明确规定，使用他人作品必须经过著作权人的许可，但因个人学习、时事报道、科学研究等十二种情形下的使用可以不经著作权人许可，也不用向其支付报酬，但应当指明作者的姓名、作品名称等。因此，图书馆在进行 AIGC 内容生成中，应对其数据库中存储数据和从图书馆外获得的数据所有权归属进行分析判断，按照数据资源使用的对象、方法、途径和获利对象进行分类管理，对受到版权保护的作品应采取合法方法获得资源的使用授权。此外，对于图书馆利用 AIGC 生成的阅读场景和内容，也应通过第三方版权认证平台利用区块链、大数据挖掘、模型算法、人工智能等前沿技术，对生成的语音、文字、图片、视频等内容，进行登记上链、侵权监测、原创比对等全链路版权服务，保护图书馆 AIGC 应用所生成的阅读场景和内容的知识产权 [20]。

2.4 对 AIGC 应用数据进行生命周期全过程的监管和保护

在图书馆 AIGC 应用中，其生成的内容并非真实的现实物理世界数据，是图书馆为了保证读者智慧阅读活动科学、高效、个性化、高收益、实时性和经济性，而虚拟的阅读场景和内容，因此，AIGC 阅读实践具有较强的应用风险和不确定性 [21]。图书馆如果不结合数据生命周期全流程而开展对 AIGC 应用数据的监管、验证和保护，将会导致读者 AIGC 阅读活动生成错误的场景和内容，进而影响读者建立正确的价值观和知识获取。因此，图书馆必须对 AIGC 应用数据进行生命周期全流程的监管和保护。

以图书馆历史场景生成训练为例，其数据

主要来源于图书馆自有数据、采购共享数据和网络爬取数据，数据内容包括网络文本、历史知识图库、音频与视频数据等[22]。《中华人民共和国个人信息保护法》第十四条规定，“个人信息的处理目的、处理方式和处理的个人信息种类发生变更的，应当重新取得个人同意。”因此，如果图书馆将这些数据运用于 AIGC 的阅读场景生成之外的模型训练，还必须获得版权拥有者的二次授权，公民有权访问、删除、更正或停止处理他们的私人数据。此外，诸如图书馆在 AIGC 机器训练中使用到海量的公民照片，也可能会因侵犯公民的面部生物识别信息而违反《中华人民共和国个人信息保护法》，必须获得照片所有者的使用授权。在对 AIGC 模型进行训练时，也要训练 AIGC 系统对敏感词语、违禁内容、AI 造假等内容的自动过滤，并在内容作品上生成 AI 标识（表明作品由 AI 生成）。

跨境数据是图书馆 AIGC 实践的一个重要数据来源。跨境数据主要由流入和流出数据两部分组成，对于流入图书馆的数据需遵守境外国家、我国和图书馆行业相关法律规定与行业规范要求，做好数据的审核、管理和使用工作[23]。对于国外用户访问我国图书馆 AIGC 应用系统而产生的数据，必须遵守《中华人民共和国网络安全法》的相关规定，关键信息基础设施的运营者在中华人民共和国境内运营中收集、产生的个人信息和重要数据应当在我国境内存储。而对于流向境外的数据，则必须依据《数据出境安全评估办法》《关于实施个人信息保护认证的公告》《网络安全标准实践指南——个人信息跨境处理活动安全认证规范 V2.0》和《个人信息出境标准合同办法》等，对图书馆 AIGC 应用风险和流出的数据进行安全评估审计后方可获得授权或许可。

2.5　对 AIGC 系统输入数据、生成算法和生成内容进行有效监管

AIGC 技术实现了图书馆读者服务模式的根本性变革，为读者提供了近似真实的虚拟还原场景，极大提升了读者的阅读收益和满意度。但是，AIGC 技术的复杂与不成熟特性，也使图书馆面临着数据真实性、生成内容合法性、漏洞病毒攻击、用户版权和伦理道德等方面的威胁与挑战。因此，必须坚持基于深度融合 AI 智能审核与人工审核相结合的原则，才能确保图书馆 AIGC 应用安全可靠[24]。

图书馆 AIGC 风险防控系统结构与检测流程如图 1 所示，主要由 AI 生成场景与内容推送层、智能风控判定流程层、生成内容风险检测层、AI 生成场景与内容层等 4 个层次组成。

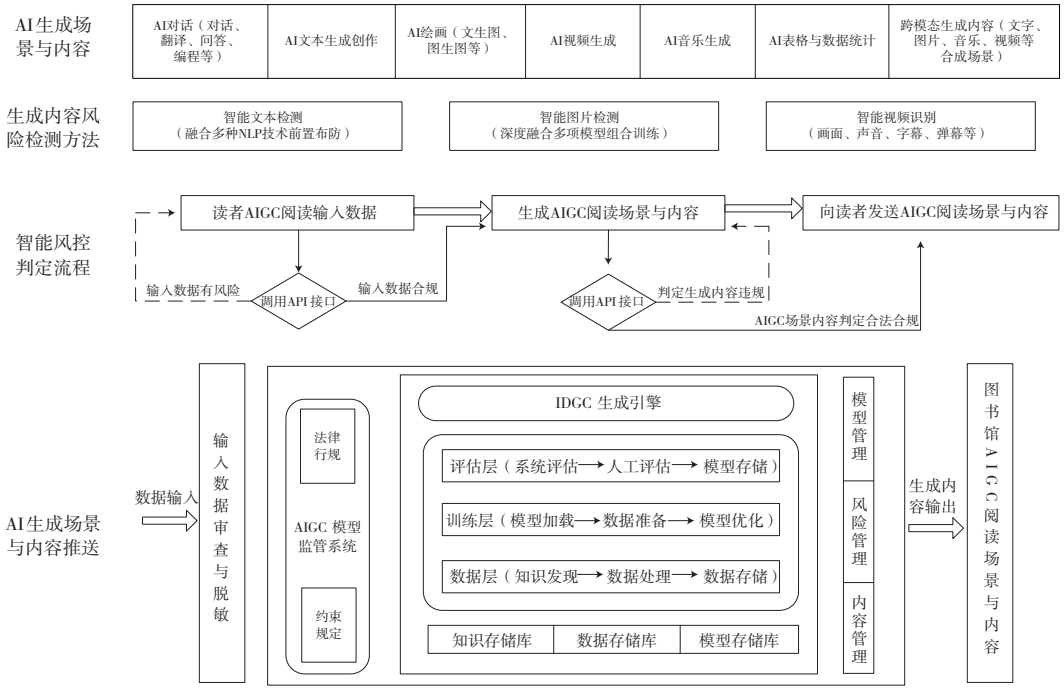

图 1　图书馆 AIGC 风险防控系统结构与检测流程图

2.5.1 AI生成场景与内容推送层

系统的最底层是AI生成场景与内容推送层，负责读者阅读活动AI场景与内容的生成、AIGC系统的构建与安全评估、读者阅读活动AIGC生成内容和场景的安全评估与推送等[25]。当系统输入端口AIGC应用初始数据后，系统首先对所输入的数据进行安全审查与敏感信息脱敏，然后将合规信息输入至AIGC生成引擎。AIGC生成引擎的最底层由知识存储库、数据存储库和模型存储库等组成，分别存储着图书馆IDGC应用所需要的行业领域知识数据、图书馆自有数据和AIGC应用模型，负责为AIGC场景与内容生成提供基础数据、AIGC应用模型的构建与训练等支持。有了底层数据存储库数据的支持，数据层可将图书馆在AIGC实践中发现的知识转化为数据并存储；训练层依据海量训练数据科学训练AIGC模型，不断对模型进行微调、评估及测试，提升模型的智能化、效率和决策正确性；而评估层由系统自动评估系统和人工评估相结合，负责对AIGC生成模型进行安全性与合法性的综合评估，着重构建以图书馆、制度和技术为主体的三位一体的安全保障体制。AIGC模型的监管系统主要以国家法律、法规和图书馆行业规定为依据，实现图书馆AIGC实践流程的模型、风险和内容管理，并将审核合格的AIGC场景与内容实时、动态地推送至读者。

2.5.2 系统智能风控判定流程层

系统的第二层是智能风控判定流程，应结合AIGC生成流程完成对生成内容输入数据、AIGC生成场景与内容的合法性判定，并将合法合规的场景与内容实时、动态推送至读者开展AIGC阅读活动。智能风控判定流程应依据国家《互联网信息服务算法推荐管理规定》和《互联网信息服务深度合成管理规定》等相关法规对AIGC应用开展安全、合规性判定，及时评估、识别、拦截、过滤和记录违规的生成合成类算法机制机理、AIGC输入数据和生成内容，并对生成内容通过标注不影响读者阅读的标识信息来明确版权。

2.5.3 生成内容风险检测层

系统的第三层是生成内容风险的检测层。系统对生成内容与场景风险的检测主要有文本、图片和视频三大应用领域，检测的内容主要涉及政治、暴力、黄色、保密、版权等内容，以及关系AIGC生成内容是否完整、准确、严谨、高价值量和符合道德规范。对智能文本的检测主要融合多种NLP（自然语言处理）技术，实现对普通及变体违规文本的识别与过滤。智能图片检测是图书馆基于海量样本图库对检测模型的持续训练与迭代更新，并深度融合Inception、ResNet、MTCNN、EAST、CRNN等多模型组合，准确识别各种违法、违规图片。结合读者AIGC阅读服务实时、动态、高数据流量的特点，智能视频的识别必须具有毫秒级系统响应时间、实时动态监测和关注视频细节的特点，并且实现对AIGC应用视频画面、声音、字幕、文本、弹幕等内容的同步监测和审查。

2.5.4 AI生成场景与内容层

系统的最高层是图书馆AIGC应用生成的场景与内容，主要由AI对话、AI文本生成创作、AI绘画、AI视频生成、AI音乐生成、AI表格与数据统计、跨模态生成内容等部分组成，是经过图书馆安全和合规检查的最终生成场景与内容，由图书馆根据读者阅读模式和用户服务途径实时、个性化智能推荐。

图书馆AIGC风险防控是一个涉及数据安全管理、算法可靠性保证和生成内容合法性治理的多内容防控，不仅要构建科学、高效的AIGC应用风险发现、判定和拦截系统，还需要建立涉及数据采集、数据存储、AIGC算法、AIGC系统模型、最终生成内容安全性判定的安全管理综合生态体系，才能实现对AIGC应用读者群、AIGC的实践全流程、AIGC应用系统的规范与监督。

3 结语

伴随着AIGC技术的快速发展和在图书馆中的广泛应用，AIGC技术重构了图书馆用户服务模式与服务应用路径。通过人工智能算法与机器自主训练，AIGC自动生成了以文本、图像和视频为主体的多模式阅读生成场景与内容，使图书馆的服务更加智能、逼真、精准、高效、多样，读者可以智慧化、个性化和创造性地开展阅读活动，确保读者知识获取流程更加实时、高效、满意和个性化。

但是，复杂、多模态的 AIGC 阅读实践活动在提升图书馆智慧服务水平和读者阅读收益的同时，也带来了极大的安全风险与隐患，虚假信息、恶意的系统软件与生成内容、病毒攻击、算法歧视等，对读者智慧阅读的合法性、公平公正性和读者隐私安全造成严重威胁。因此，图书馆在利用人工智能技术构筑 AIGC 安全应用、防护系统的同时，还必须坚持多元共治、关注数据处理与信息互动安全、重视管理与技术并举的原则，结合图书馆 AIGC 实践中数据生命周期管理流程对 AIGC 系统输入数据、AIGC 模型合成、生成的场景与内容、读者 AIGC 阅读流程等进行全面监管，才能为读者提供安全、高效、个性化和可扩展的 AIGC 智慧阅读服务。

参考文献：

[1] 中国信通院，京东探索研究院 . 人工智能生成内容 (AIGC) 白皮书（2022 年）[EB/OL]. [2024–05–01]. http://www.caict.ac.cn/kxyj/qwfb/bps/202209/t20220902_408420.htm.

[2] 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.

[3] 杜雨，张孜铭 .AIGC: 智能创作时代 [M]. 北京：中译出版社，2023.

[4] 朱荣荣 . 类 Chat GPT 生成式人工智能对个人信息保护的挑战及应对 [J]. 重庆大学学报 (社会科学版)，2024(5):1–14.

[5] 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(4):34–43.

[6] 郭如愿 . 论人工智能生成内容的信息权保护 [J]. 知识产权，2020(2):48–57.

[7] PANDA S,KAUR N. Exploring the viability of ChatGPT as an alternative to traditional chatbot systems in library and information centers[J].Library Hi Tech News, 2023, 40(3):22–25.

[8] 刘亚丽，范逢春 .ChatGPT–AIGC 用户风险感知维度识别与治理研究：基于扎根理论的探索性分析 [J/OL]. 情报理论与实践 :1–13[2024–01–22]. http://kns.cnki.net/kcms/detail/11.1762.G3.20231031.1540.004.html.

[9] 人工智能伦理问题建议书 [EB/OL].[2024–02–02]. https://www.thepaper.cn/newsDetail_forward_15596087.

[10] CHRISTOPHER C,ELIAS T.ChatGPT implications for academic libraries[EB/OL].[2024–05–04].https://crln.acrl.org/in–dex.php/crlnews/article/vies/25821/33770.

[11] European Commission.DigComp2.2:The Digital Competence Framework for Citizens–with New Examples of Knowledge, Skills and Attitudes[EB/OL].[2024–05–04].https://op.europa.eu/en/publication–detail/–/publication/50c53c0–abeb–11ec–83e1–01aa75ed71a1/language–en/format–PDF/source–268854425,2024–05–08.

[12] 陈永伟 . 超越 Chat GPT：生成式 AI 的机遇、风险与挑战 [J]. 山东大学学报 (哲学社会科学版),2023(3):127–143.

[13] 蔡士林，杨磊 .Chat GPT 智能机器人应用的风险与协同治理研究 [J]. 情报理论与实践，2023, 46(5):14–22.

[14] 搜狐 .AIGC/AI 生成内容产业展望报告 [EB/OL]. [2024–05–11].https://www.sohu.com/a/639161593_121615303.

[15] 量子位 .AIGC/AI 生成内容产业展望报告 [EB/OL].[2024–05–11].https://www.djyanbao.com/report/detail?id=3425715 & from=search_list.

[16] 国家网信办等七部门联合公布《生成式人工智能服务管理暂行办法》[EB/OL].[2024–05–11]. http://www.cac.gov.cn/2023–07/13/c_1690898326795531.htm.

[17] 刘智锋，吴亚平，王继民 . 人工智能生成内容技术对知识生产与传播的影响 [J]. 情报杂志，2023, 42(7):123–130.

[18] 李白杨，白云，詹希旎，等 . 人工智能生成内容 (AIGC) 的技术特征与形态演进 [J]. 图书情报知识，2023(1):66–74.

[19] 蒲清平，向往 . 生成式人工智能：Chat GPT 的变革影响、风险挑战及应对策略 [J]. 重庆大学学报 (社会科学版), 2023,29(3):102–114.

[20] LUND B D，WANG T. Chatting about ChatGPT: how may AI and GPT impact academia and libraries? Library Hi Tech News. [EB/OL].[2024–05–19].https://

www.emerald.com/insight/content/doi/10.1108/LHTN-01-2023-0009.

[21] 丁波涛，夏蓓丽，范佳佳. 全球信息社会蓝皮书：全球信息社会发展报告(2022)[M]. 北京：社会科学文献出版社，2022:1-37.

[22] 2023年AICC发展趋势报告[EB/OL].[2024-05-17]. https://www.199it.com/archives/1558601.html.

[23] OPENAI. ChatGPT[EB/OL].[2024-05-18]. https://chat.openai.com.

[24] 叶鹰，朱秀珠，魏雪迎，等. 从Chat GPT爆发到GPT技术革命的启示[J]. 情报理论与实践，2023,46(6):33-37.

[25] 李白杨，白云，詹希旎，等. 人工智能生成内容（AIGC）的技术特征与形态演进[J]. 图书情报知识，2023,40(1):66-74.

马晓亭 兰州财经大学信息工程与人工智能学院。甘肃兰州，730020。

（收稿日期：2024-06-12 编校：曹晓文，左静远）

江苏省图书馆学会编译出版专业委员会召开成立大会

2024年11月17日，江苏省图书馆学会编译出版专业委员会的成立大会在镇江隆重召开。此次盛会由江苏省图书馆学会主办，江苏大学图书馆与南京大学出版研究院携手承办。大会汇聚了来自国家图书馆、南京图书馆、南京大学图书馆、苏州大学图书馆、泰州市图书馆等近40家图书馆领导，以及国家图书馆出版社、南京大学出版研究院等多家出版社和科研机构领导同仁，大家集聚智慧，共同探讨我省图书馆编译出版领域的未来发展。

江苏省图书馆学会编译出版专业委员会主任、南京大学出版研究院副院长、国家新闻出版署智慧出版与知识服务重点实验室主任杨海平教授主持开幕式，与会的各位领导先后致辞，对编译出版专业委员会的成立表示诚挚的祝贺。会上，江苏省图书馆学会李浩秘书长宣读了专业委员会的委员名单，并颁发聘书，对委员会未来的工作寄予厚望。

在专家主旨报告环节，中国期刊协会杨树弘副会长以《构建适应全媒体生产传播的工作机制和评价体系》为题，就全媒体时代的核心价值，分析探讨了主流媒体构建深度融合的路径及科学的评价体系。《中国图书馆学报》吴澍时常务副主编则带来了《中国特色图书情报学专栏建设思考及展望》，期望通过专栏建设推动中国图书情报学的持续发展与国际交流。南京大学图书馆副馆长邵波教授、南开大学教授徐建华、南京师范大学图书馆馆长姜晓云教授、南京农业大学信息管理学院党委书记郑德俊教授等多位专家，分别围绕智慧图书馆建设、电子书新书馆配、信息化时代的阅读策略及智慧、与用户共创知识服务价值等前沿议题发表了精彩演讲。中国知网数字出版中心总经理谢磊详细介绍了知网大模型在出版业与学术科研领域的应用，江苏大学科技信息研究所所长刘桂锋则对图书情报学术期刊与学科建设的互融共生关系进行了深入剖析，为相关领域的发展提供了创新启示。

成立大会结束后，江苏省图书馆学会编译出版专业委员会随即召开工作研讨会，由泰州市图书馆馆长乔立兵主持。杨海平教授说明了专委会的工作目标，即促进江苏省信息资源管理学术研究成果的出版、发行与评价，鼓励和助力江苏省信资管研究者和工作者学术成果的发表。委员们从江苏省信息资源管理学科优秀学术成果编辑出版发表工作、优秀研究者和工作者奖励、新时代图书馆资源建设与馆配转型、与中图学会编译委同步共振、AI助力论文写作与发表等多方面共同规划专委会的工作重点及落实措施。江苏省图书馆学会编译出版专业委员会的正式成立，将为推动江苏省乃至全国信息资源管理领域的学术研究与发展注入新的活力与动力。

（编译出版专业委员会供稿）