DOI:10.16384/j.cnki.lwas.2024.10.004

智慧图书馆领域生成式人工智能技术应用的法律容错机制*

童云峰　张学彬

摘　要　文章以ChatGPT为例，分析智慧图书馆领域中生成式人工智能技术应用面临的技术风险和法律风险，提出需构建图书馆领域生成式人工智能技术应用的法律容错机制，即广泛适用第一位阶的自律规则，防控技术风险；主要适用第二位阶的前置法律，规制应用风险；渐进适用第三位阶的刑事法律，轻罪优先于重罪。

关键词　智慧图书馆；生成式人工智能；ChatGPT；法律容错机制；比例原则

分类号　G250.76

本文引用格式

童云峰，张学彬.智慧图书馆领域生成式人工智能技术应用的法律容错机制[J].图书馆工作与研究，2024(10)：23-32.

1　引言

2022年11月，美国OpenAI公司推出的基于GPT-3.5架构的ChatGPT引起社会各界广泛关注；2023年3月，GPT-4的升级换代再次引发人工智能技术研究热潮。图书馆领域学者关注生成式人工智能技术对图书馆发展的影响，认识到其在管理方案制定、文献知识获取等方面拥有广阔的应用前景[1]。同时，现有研究也关注该技术应用于图书馆建设过程中存在的风险，但提出的应对方案多立足管理学或公共政策视角[2]，鲜少基于法律视角。2023年8月15日起施行的《生成式人工智能服务管理暂行办法》[3]（下文简称《管理办法》）提出针对生成式人工智能技术的多元治理模式，可为图书馆领域应对生成式人工智能技术应用风险提供基本规范。而该规范如何具体适用于图书馆领域，保持图书馆领域生成式人工智能技术应用发展创新与风险防控的平衡，成为学界和业界面临的重要问题。鉴于此，本研究以ChatGPT为例，分析智慧图书馆场景中生成式人工智能技术应用风险，论证图书馆领域构建生成式人工智能技术应用法律容错机制的可行性，并提出该机制的贯彻路径，以期充分激发生成式人工智能技术的创新活力，切实推进智慧图书馆建设。

2　生成式人工智能赋能智慧图书馆建设的可行性与伴生风险

智慧图书馆建设需要智能技术的加持，生成式人工智能技术与图书馆智慧化建设具有同源性。因此，准确识别生成式人工智能技术应用风险对智能图书馆建设至关重要。

2.1　生成式人工智能应用于智慧图书馆建设的可行性

数字时代，用户对图书馆智能化服务的要求越来越高。生成式人工智能以人为本，具有人机交互能力，能最大程度契合用户的语言模

*本文系上海市哲学社会科学规划项目青年课题“数字时代个人信息权益的全生命周期刑法保护研究”（课题编号：2023EFX010）研究成果之一。

收稿日期：2023-12-20　　**本文编校**：张雪英

式、满足用户的具体需求，因而生成式人工智能应用于智慧图书馆建设具有现实可行性。

2.1.1 生成式人工智能与智慧图书馆的高度关联性

生成式人工智能的三“大”典型特征与智慧图书馆建设的底层逻辑相契合。具体而言：

其一，大数据。生成式人工智能的知识广度以分析与转化的海量数据为基础。OpenAI推出的初代自然语言处理模型GPT-1的训练参数量为1.17亿，革新后的GPT-3使用的训练参数量高达1750亿[4]。而目前ChatGPT已经升级为GPT-4，虽然研究人员尚未公布其训练参数量，但根据其强大的文本生成能力可以推测，GPT-4可能拥有近万亿级的数据库。由此可知，大数据是生成式人工智能产品迭代升级的基础。

其二，大算力。生成式人工智能依托Transformer神经网络架构，采用“自回归”模式对输入文本进行“预测”。ChatGPT搭载人类反馈强化学习(Reinforcement Learning from Human Feedback，RLHF)、指示微调(Instruction Tuning)、思维链(Chain of Thought)等技术，同时依赖超级计算机、GPU芯片等软、硬件设施，能在用户输入文本几秒钟内捕捉、领会用户意图，进而输出质量高、逻辑性强、富有人性化的文本，这说明其具有强大的算力系统。

其三，大算法。算法是生成式人工智能的“大脑”，从语料库的标注与训练到各类技术的组合与运用，再到最终文本的推理与输出，均有赖于高质量的算法。算法具有自主性，能在无人工监督的情况下协调各类组成单元，“指挥”ChatGPT的用户需求识别与分析工作，根据不同需求准确生成相应文本。但ChatGPT仍为弱人工智能产品，其仅能按照设计和编制的程序运行并实现人类意志[5]。

由此可见，具备上述三大特性的ChatGPT类生成式人工智能产品本质上是一种技术工具，其所展现的强大的数据集合和分析能力与图书馆智能化建设需求相契合。具体而言，进入数字时代，数字资源日趋成为文献资源的主要形态，而ChatGPT是推进智慧图书馆建设的重要动力。在数据层面，图书馆服务始终建立在对文献资源、浏览记录、运行参数、政策信息等数据的存储、识别与分析基础上，并通过构建一体化描述模型服务用户[6]。ChatGPT类生成式人工智能能存储上述数据，通过分析输入数据实现文本输出，满足用户需求。在技术层面，智能技术是智慧图书馆的核心驱动力[7]，智能文献咨询、智能系统维护、智能设施管理等是其建设方向。而ChatGPT集合了智慧图书馆建设运行所需的数据整合、统计分析与深度学习技术，搭载了强大的算力系统与算法规则，可处理图书馆数据，继而满足用户检索需求，为图书馆管理方案优化提供支持。

2.1.2 智慧图书馆场景中生成式人工智能应用前瞻

图书馆服务存在服务环境、数据、信息主体之间进行信息流转的链式依存关系[8]，故基础设施、数据资源和服务应用三要素的优化组合成为构建多层次图书馆智慧信息服务体系的主要方向[9]，而生成式人工智能的应用可在其中产生积极作用。具体而言：

其一，提升图书馆技术联结能力，促进图书馆基础设施迭代更新。图书馆提供智慧信息服务过程中需依托智能技术对数据进行采集、挖掘与分析，而生成式人工智能本身蕴含强大的算力、算法系统，在一定程度上是对机器学习、大数据分析、自然语言处理、可视化分析等智能技术的深度联结。生成式人工智能的规模化应用不仅可提高智能技术的集合使用效率，而且可释放依托该智能技术运行的软、硬件设施的使用潜能，促进图书馆对传感器、监测器、门禁系统、操作系统等高性能基础设施的应用，满足图书馆智能化运行的需要。

其二，整合分析图书馆数据资源，助力解决数据壁垒、数据孤岛等问题。图书馆数据主要有3种类型：①管理数据，如图书馆基础设施

的运行数据、后台控制的计算机系统数据等；②资源数据，如馆藏或馆际交互的文献数据、各类数据库与知识库的应用数据等；③用户数据，如浏览记录、收藏、评论、借阅等行为数据等[10]。可见，图书馆数据种类繁杂、形式多样、体量巨大，且表达方式不一，数据整合分析难度较大。而生成式人工智能可通过自身强大的深度学习能力分析各类数据库的共性特征及各类数据的内在联系。虽然生成式人工智能无法提供可视化的数据分析过程，但其最终能通过符合人类习惯的自然语言输出可用结果，对促进数据共享、挖掘数据价值、推动资源互联互通等具有积极意义。

其三，为用户、馆员分别提供个性化服务方案，满足其需求。ChatGPT 已实现插件支持[11]，其可接入图书馆系统，实现数据管理。具言之，对馆员而言：①实现决策支持。如生成式人工智能可分析用户的浏览记录、笔记收藏、互动内容等，帮助馆员掌握文献存储现状与用户阅读偏好。这不仅能辅助馆员制订选书、购书计划，还可及时调整纸质文献布局，实现资源优化配置。②实现程序设计。智能选座、购书、门禁等系统的程序需由专业人员编写，生成式人工智能掌握多种编程语言的结构与程序编写规则，具有撰写脚本和代码的能力。专业人员输入相关要求，ChatGPT 会直接给出源代码[12]，帮助其查验、完善程序编写工作。对用户而言，生成式人工智能可促进用户的知识获取方式从“检索提取”向“一站式查询”转变。具言之：①辅助阅读。生成式人工智能可根据用户阅读方向、目标和计划向其推荐书目，并告知查阅方式。当用户在阅读中遇到疑难问题时，生成式人工智能可根据用户的多轮提问提供情景式解答。②辅助写作。生成式人工智能可借助馆藏文献系统分析文章选题和框架，帮助用户了解发文趋势，提供实例和写作思路，推荐可用的文献资源。写作完成后，其也可润色文章语言、评估文章质量，为个性化写作提供支持[13]。③便捷使用其他服务。用户可通过生成式人工智能预先了解图书馆内阅览坐席使用和基础设施设置情况。此外，随着馆际交流的频繁和深入，生成式人工智能可掌握多家图书馆大数据，进一步提升图书馆服务能力。

2.2　智慧图书馆场景中生成式人工智能规模化应用的双重风险

虽然生成式人工智能的应用能够助力智慧图书馆建设，但其也蕴含着两类风险，即因初创技术漏洞产生的技术风险和系统应用过程中产生的法律风险。通过社会系统理论的范式分析可以确知，这两类风险既各自独立又相互联系。一方面，两类风险属于不同的话语体系，技术风险是内生性风险，法律风险需要法律的检视与规制，是评价体系视角下的风险类型；另一方面，两类风险具有一定的关联性，有些技术风险会衍变为法律风险，继而受到法律规制，而法律风险的化解又需要基于技术理性，不能盲目动用国家权力。

2.2.1　天然性技术风险

生成式人工智能在不同运行阶段可能产生不同的技术风险，主要包括 3 个方面：

其一，运行设施方面的风险。运行生成式人工智能需要强大的软、硬件设施，其与图书馆内部系统的连接涉及计算机运行、存储等核心技术和资源的引进与使用。生成式人工智能发展初期，图书馆可能面临资金成本巨大、技术“卡脖子”等市场风险[14]。此外，网络安全也会加剧技术风险。目前生成式人工智能仍处于发展初期，其自身防护措施及避免接口端传导风险的技术防护措施尚未健全，图书馆内部系统一旦遭到不法分子入侵，依靠生成式人工智能技术组成的智能系统也将受到攻击，进而引发系统性风险。

其二，输出过程方面的风险。生成式人工智能存在遮蔽数据处理和文本输出过程的“算法黑箱”[15]。算法具有 3 种特性：①价值偏向性。生成式人工智能的数据选择、规则设计体现开发者的主观价值倾向。若图书馆、上游供

货商或开发人员通过设置不公平的算法规则，使某一文献资源频繁被优先推荐，将破坏图书馆服务的公平性，损害用户信息知情权等个人权益。②不透明性。算法多涉及商业秘密，即使对外公开设计的基础代码，普通人也很难理解。该特性将导致文本输出过程无法被实时监控，继而导致监管事后化，给违法犯罪留下可乘之机。③使用必要性。算法是生成式人工智能运行不可或缺的基本逻辑和核心规则，但其也提高了风险发生的概率。

其三，输出内容方面的风险。由于生成式人工智能主要使用以文本为中心的语料库进行训练，故其无法实现复杂的数理运算和逻辑推理。虽然使用深度学习技术能使其输出复杂文本，但其依然无法理解数据或文本之间的逻辑关系与知识关联，极易生成不合理内容。若其为馆员输出此类不合理代码程序，在其无法应用或造成系统错误时，需要馆员反复检查和纠错，进而导致管理效率低下；若其为用户输出此类内容，可能误导用户，甚至引发学术造假。

2.2.2 伴随性法律风险

生成式人工智能的技术联结蕴含一定的外溢风险，技术风险可能衍变为法律风险，主要表现为3类：

其一，知识产权侵权风险。笔者认为，生成式人工智能的生成物可成为知识产权保护的客体，使用者应在一定范围内享有知识产权。以著作权为例，在保护客体认定方面，ChatGPT生成的内容符合“独创性”标准。著作权只保护具有独创性的表达，不保护创作过程或通过创作而体现的思想，ChatGPT可以生成具有创造性“表达”的内容，且其表达方式“千变万化”。因此，不能因ChatGPT不具有人类的“思维或方法”而否认其生成物的独创性。在权利主体认定方面，依据实质贡献原则与利益衡量原则，ChatGPT的开发者享有ChatGPT生成内容的知识产权，但特别约定的除外[16]。该观点在腾讯公司诉“网贷之家”一案中有所体现[17]。然而，ChatGPT的使用条款指出，OpenAI会将输出内容的所有权转让给输入内容的用户[18]。这不仅符合民法意思自治的法治原则，而且肯认了使用者的权利主体地位，可充分发挥知识产权制度对使用者的激励作用。若将知识产权赋权给开发者，将极大限制对生成式人工智能生成物的利用，继而无法合理分配因生成物的商业利用产生的利益，人工智能产业发展将会受挫[19]。

具体而言，生成式人工智能的应用将产生两类著作权侵权风险：①生成作品时的风险。生成式人工智能使用的数据库体量巨大，其中不乏受著作权法保护的内容。生成式人工智能帮助用户形成论文大纲和写作思路时可能会对该内容进行词语重组、语义关联、识别转换，继而形成逻辑相似的内容，存在侵犯著作权权利人的修改权、复制权和改编权的风险。②利用作品时的风险。当用户将ChatGPT生成的读后感署名并上传至网络时可能侵犯著作权权利人的署名权和信息网络传播权。在不满足合理使用且符合侵权构成要件的前提下，用户对ChatGPT生成的外文文献的翻译与汇编也可能引发翻译权、汇编权侵权风险。总之，生成式人工智能应用产生的著作权风险根源于数据库使用，也与算法的自主学习与分析能力相关。

其二，数据利用与内容无序生成风险[20]。一方面，生成式人工智能将用户输入的信息作为自身学习的基础数据，并刻画出用户画像，帮助馆员掌握用户阅读行为及偏好，提供精准推送服务，侵犯了用户的人格利益[21]。而生成式人工智能的隐私政策或使用说明中规定的获取用户信息的方式、种类及获取信息的使用范围与权限等是否符合法律法规规定及能否贯彻落实有待观察。此外，若将生成式人工智能与图书馆内部系统相连接，一旦系统遭到非法攻击，将产生危及个人隐私、泄露商业秘密的法律风险[22]。另一方面，由于生成式人工智能数据库可能存有欺诈性、误导性信息，故其

可能生成欺诈性、误导性内容[23]。如因存在系统漏洞，其可能在用户诱导下对“如何不用归还借书”作出解答。这一特性若被不法分子利用，将加剧侵犯图书馆合法权益的法律风险。

其三，算法衍生的法律责任归责风险。生成欺诈、虚假信息或获取、分析个人信息均以算法规则为基础，故前述法律风险均为算法引发的风险。总体而言，ChatGPT 不能作为责任主体，由算法运行产生的损害责任应由其开发者或管理者承担。而在图书馆场景下，该归责问题具有一定的特殊性。在归责主体方面，生成式人工智能的应用涉及开发人员、馆员、基础设施维护人员、用户等多方主体，其均可对算法运行产生影响，如何在具体情景下协调责任分配值得研究。在归责认定方面，既有法律规范如何应对算法风险对因果关系、证明责任等的异化与冲击值得关注。在归责限度方面，法律对算法或相关人员行为的规制会在一定程度上约束技术的发展。因此，法律应在风险规避、技术创新和权益保护之间寻求平衡[24]。

综上，生成式人工智能的应用能够促进图书馆智慧化建设，但也应关注其应用产生风险的多样性和系统性。而如何保持风险防控与技术创新的平衡，促使生成式人工智能在图书馆领域发挥最大效益成为亟需学界与业界探讨的问题。

3　智慧图书馆领域生成式人工智能技术应用法律容错机制的提出

生成式人工智能技术应用可能产生较为严重的法益侵害结果，需要法律规制。然而其为初创技术，若法律规制手段过于严苛将导致开发者和使用者动辄得咎，从而阻碍技术创新与应用。因此，需塑造法律容错机制，即法律对智慧图书馆领域生成式人工智能技术应用持适度宽容态度，预设技术适度犯错空间，当技术应用行为超越必要限度时适用法律规制，尤其是刑法规制。对此，需遵循“自律法治→前置法治→刑事法治”的图书馆法治递进方案[25]。

3.1　智慧图书馆领域自律规则的广泛适用

相较于法律，自律规则的惩戒力度较轻，宽容度较高。因此，应广泛和优先适用自律规则，超越自律规则调适范围的行为才有必要适用法律规范，为智慧图书馆领域生成式人工智能技术应用创造相对广阔的自由空间。

一方面，“代码即法律”原则说明代码等技术规则可部分承担法律规制作用。技术规则虽不如法律规范严厉，但其依然发挥指引、约束和调适功能。其一，技术规则由程序员通过代码形式表达[26]，具有编写主体非国家性、具体内容专断性的特点，故开发者和馆员对生成式人工智能技术的技术规则最为熟悉，其普适性较强，激励“代码自治”虽不能根除风险，但能起到一定的防范作用。其二，生成式人工智能技术应用风险主要发生在图书馆的数字空间，现实的法律不能完全适用，或存在法律适用鸿沟。而技术规则生成并用于控制生成式人工智能技术学习、分析与输出的内容及其过程，不存在适用鸿沟。

另一方面，图书馆行业规则具有“软法性”，其对生成式人工智能技术应用较为宽容。在图书馆等专业领域中，根据自身特点及生成式人工智能的使用情况而制定的行业规则更具针对性、灵活性，是对法律规范的有力补充[27]。充分发挥行业规则的规范作用不仅可为技术发展创造广阔空间，还可节约法律成本。虽然行业规则对开发者、馆员、用户等主体的行为并无强制约束力，但对违规行为有一定的惩戒作用，其中的惩戒措施既能收到“点到为止”的效果，也不会挫败开发者和使用者的积极性。换言之，图书馆行业规则的广泛适用可为生成式人工智能技术应用创造自律规则与法律规范之间的容错空间。

3.2　智慧图书馆领域前置法律的主要适用

当自律规则无法调适生成式人工智能技术应用中的权益损害行为时，需要法律介入。

当法律适用者对其是否适用刑法而犹豫时，若优先适用民法、行政法等前置法，也是在塑造容错机制。相较于刑法，前置法处罚措施较轻，不会使行为人面临“自由刑”“生命刑”和“前科”的惩罚，对开发者和使用者积极性的挫败程度也较低。同时，前置法也能在一定程度上肩负防控风险的重任。

一方面，民事法律对生成式人工智能技术应用风险具有较好的规制效果。其一，民事法律能够规制侵权行为并救济法益。如当 ChatGPT 过度分析与利用用户个人信息、侵犯用户个人信息权益时，用户可提起民事诉讼，用户群体、组织可提起公益诉讼，以维护自身权益。通过民事诉讼提供救济启动迅速、便捷，能较快恢复被侵害法益。其二，出于利益衡量，掌握数据资源主动权的图书馆有时会以侵犯用户权益为代价换取更为优质的服务资源。而民事法律的经济制裁和赔偿损失惩罚手段能发挥威慑作用，有效防止行为人再次制造风险，可在一定程度上降低相对严重的技术应用风险与权益损害。

另一方面，行政法对生成式人工智能技术应用风险具有一定的威慑效果。其一，行政法在防控风险的同时也注重法益保护。如为避免 ChatGPT 生成侵犯知识产权的内容，行政法可能会限制其使用某些图书馆数据库，这不仅可以防控风险，还可以保护相关作品的知识产权。其二，行政法可以更好地保护行为主体开发和应用生成式人工智能技术的积极性。如生成式人工智能技术系统一旦受到攻击，将泄露大量个人信息。若对实施攻击行为的不法分子适用行政法和刑法均涉及限制自由或财产罚，规制效果具有重合性；若适用刑法会给行为人带来“前科”等附随处罚效果，或将阻碍人们对生成式人工智能等新兴技术的探索；若适用行政法，则不会挫伤相关行为主体研究、开发和使用新技术的积极性，为图书馆领域生成式人工智能技术的应用创造容错空间。

3.3 智慧图书馆领域刑事法律的渐进适用

治理图书馆领域生成式人工智能技术应用风险，刑法必不可少。但刑罚是最严厉的制裁措施，若运用不当会严重阻碍技术的应用和发展。因此，刑法介入应谨慎，非必要不适用刑法，能适用轻罪则不适用重罪，从而在刑法内部为图书馆领域创造生成式人工智能技术应用的容错空间。

一方面，被生成式人工智能技术应用行为严重侵犯的法益需要刑法修复。具体包括以下法益：①知识产权。若用户恶意利用生成式人工智能技术批量生成具有著作权争议的内容且上传至网络获取经济利益，可能构成侵犯著作权罪。②一般人格利益与隐私权。与传统平台相比，利用生成式人工智能技术收集的用户信息体量更大，刻画的用户画像更精准，管理者过度收集、非法使用该信息将构成侵犯公民个人信息罪[28]。③因算法而侵犯的权益。如馆员通过算法向用户推荐特定文献，可能侵犯用户享受信息自由的平等权。若馆员将自身价值观念引入算法，可能导致生成式人工智能输出具有偏见性的内容，继而区别对待不同读者，造成对用户人格权的严重侵犯[29]。只有上述行为造成严重法益侵害结果，如使被害人遭受电信诈骗，才有必要适用刑法。

另一方面，刑事合规是生成式人工智能技术应用与智慧图书馆建设的共同诉求。生成式人工智能技术的联结性及其产生风险的系统性催生了多元共治的治理模式，当其他规范不足以应对严重风险时，方可求助于刑法，此时刑法成为“刚需”。无论是为馆员设定管理职责，还是为用户提供救济措施，均离不开刑法的保障性作用。有学者认为，有必要将自我保护技术、企业内部制裁制度等刑事合规计划引入人工智能评价机制，以规避人工智能风险[30]。实际上，围绕生成式人工智能风险治理的一系列措施便是图书馆场景下的“刑事合规计划”，只有通过刑法的实施与引领，才能保障其他规范不缺位、不越位，实现预防犯罪、消解风险的规制效果。

总之，在面对生成式人工智能技术应用风险时，具有保障性的刑法在多元共治模式中不可或缺，但刑法规制并非毫无节制，需明确介入限度，以充分形塑图书馆领域生成式人工智能技术应用的法律容错机制。

4　智慧图书馆领域生成式人工智能技术应用法律容错机制的贯彻

对于智慧图书馆领域生成式人工智能技术应用法律容错机制的贯彻，笔者认为，可秉持法益保护原则和比例原则：先通过自律规则处理技术风险，尊重生成式人工智能的技术逻辑；当技术自治失效时，相继穷尽民法、行政法规范，保护法益；当前置法失效时，介入刑法。

4.1　广泛适用第一位阶的自律规则，防控技术风险

我国《互联网信息服务深度合成管理规定》[31]第5条规定，鼓励相关行业加强行业自律，建立健全行业准则，督促深度合成服务提供者和技术支持者依法开展业务。具言之，自律规则应以法律规范为依据，设置相应的技术措施与技术规范，以落实第一位阶的法律容错机制。

首先，加强技术研发，设置防控措施，防范技术漏洞。以“技术”防控“技术”是应对生成式人工智能技术应用风险的首选措施。该类风险防控措施能及时弥补技术漏洞，阻滞内生风险的发生与扩大，延缓法律介入。如可开发自动识别风险的技术程序，当图书馆内部系统遭到入侵时，及时阻断内部系统与生成式人工智能的连接，以保护用户个人信息和相关数据。当然，技术规则的自治并非绝对安全可靠，其只能为风险防控提供技术方案，自治一旦失灵则需诉诸法律。

其次，行业规则及其具体细则可作为适度免责的依据。图书馆行业学（协）会或图书馆可根据我国《民法典》《个人信息保护法》《公共图书馆法》及个人信息安全相关国家技术标准，制定关涉生成式人工智能获取个人信息的方式、权限、内容，以及用户享有的权利等行业规则，这也是《管理办法》意图重点解决的问题。用户在使用ChatGPT等服务时，图书馆应告知其输入个人信息可能面临的风险，并取得用户单独、明确的同意。此时，该行业规范即为非标准化“合同”，用户的阅览、知悉与同意使生成式人工智能处理个人信息变得合理合法，故其不涉及法律责任，无需法律介入。

最后，图书馆内部章程亦可成为矫治轻微风险的规范。除技术规则和行业规则，图书馆也会制定自治章程，规定生成式人工智能技术应用的场景和范围，用以约束图书馆、馆员、用户等主体的相关行为。数字时代，图书馆应遵循法治原则，根据《民法典》《个人信息保护法》《公共图书馆法》等法律规范制定契合时代需求的管理章程，凸显对生成式人工智能技术应用行为的规制，规避潜在风险和危害。

4.2　主要适用第二位阶的前置法律，规避应用风险

诸多风险可通过前置法进行治理，能够通过前置法保护的法益无需刑法介入，这是在法律领域内部为生成式人工智能技术应用构建第二位阶的容错机制，优先适用前置法体现法律对初创性生成式人工智能技术的宽容态度。

首先，穷尽民法规范应对生成式人工智能技术应用风险。一方面，以意思自治为基础形成的协议可作为阻却刑事违法性的事由。如Open AI官网以在线隐私声明的形式对ChatGPT生成内容的所有权益作出安排。图书馆也可将生成式人工智能技术在本馆应用时产生的风险以在线服务协议的形式告知用户，一旦用户表示确认并同意，即视为协议成立并生效。即使违背协议内容产生纠纷，仅需通过合同法或侵权责任法予以解决，无需过早适用刑法。另一方面，馆员履行管理义务时应阻却刑事违法性。以算法为例，若馆员对生成式人工智能技术应用可能产生的风险具有足够预判，强化对算法的道德性、规范性训练，设置相应

的反内容歧视和虚假机制，并尽到相应的注意义务时，ChatGPT 仍生成侵犯用户权益的内容，那么不应当要求馆员承担刑事责任，馆员只需根据侵权责任法或合同法之规定承担合理的赔偿或补偿责任。

其次，穷尽私法后的生成式人工智能技术应用风险治理应由行政法调控。国务院发布的《新一代人工智能发展规划》[32]提出应建立健全人工智能监管体系。为革除传统监管方式信息不对称的弊端，智慧图书馆时代的行政监管应走向科技监管，建立实时性和敏捷性的监管机制。该机制能有效预警风险，提升风险应对效力。因此，只要不涉及个人法益严重侵害行为，均可由该监管机制进行识别、控制与管理。如馆员恶意利用 ChatGPT 生成某一馆藏图书的暴力性内容，继而引发众多用户复制、转发、点赞、评论。虽然其引发了一定的讨论，但尚未对现实生活秩序造成严重影响。因此，此类行为通常不会直接侵犯法益，仅是单纯违反国家行政管理规范[33]，故只需行政监管部门及时预警并阻断信息传播，由行为人承担行政责任，无需适用刑法。

4.3 渐进适用第三位阶的刑事法律，轻罪优先于重罪

当穷尽前述规范仍无法化解图书馆领域生成式人工智能技术应用风险时，具有保障性作用的刑法应及时补位。但基于比例原则与刑法谦抑主义，刑法规制也应谨慎而行，遵循轻罪优先于重罪原则。如此方可在刑法内部为生成式人工智能技术应用构建容错机制。

首先，对通过增设新罪规制生成式人工智能技术应用风险持审慎态度。有学者认为，可在刑法中严格确立人工智能产品研发者和使用者的责任，增设人工智能事故罪[34]。在强人工智能时代，该罪或有适用空间。然而，比例原则之必要性原则要求须采取能够实现正当目的且相对缓和的措施。对生成式人工智能技术应用风险的治理而言，显然并无增设新罪的必要。因为该风险可能涉及个人信息权益、知识产权、社会秩序、社会安全等法益，故设立这种极具包容性的罪名会导致保护法益泛化，造成既有刑法规范不协调[35]。换言之，新科技时代并未肇生全新的法学问题[36]，生成式人工智能技术应用风险治理完全可以适用既有刑法规范中的罪名，而无需增设新罪[37]。

其次，对生成式人工智能技术应用行为可能触犯罪名的构成要件作缩小解释。根据比例原则的均衡性原则要求，对利益的限制越大，所要实现的目的须越重要。如在认定非法获取计算机信息系统数据罪时，对智慧图书馆建设需求的限制与对应用数据与网络管理秩序的保护需达到合理平衡。换言之，为促进技术创新发展，须对相关罪名的构成要件作缩小解释，防止相关罪名过度适用。如应将“违返国家规定”严格限制为我国《刑法》第 96 条规定的规范，不能包含部门规章等低层级的规范[38]93。应将所获取数据类型严格限制在具有使用意义的身份认证信息，内容失效且不会侵犯法益的信息不宜纳入保护范围[38]94。

最后，当生成式人工智能技术应用行为可能同时触犯轻罪与重罪时，应在能够保护法益的前提下优先适用轻罪。如馆员恶意利用 ChatGPT 诱导用户输入商业秘密，但因多种因素无法证明获取信息为商业秘密时，可根据行为人获取的用户其他信息的数量或违法所得，适用侵犯公民个人信息罪。若无法确定个人信息的类别，也可根据其获取的后台数据的数量与违法所得，适用非法获取计算机信息系统数据罪。应注意的是，罪名确定应符合相关罪名的构成要件，即在遵循比例原则适当性要求下保持刑法谦抑姿态。简言之，在同时触犯轻罪与重罪时，应优先适用轻罪，只有轻罪无法周延保护法益时，才应适用重罪，这是在刑法内部构建法律容错机制的应有之义。

5 结语

处于探索阶段的生成式人工智能技术发

展迅速并应用于图书馆领域，法律尤其是刑法在面对其引发的潜在风险时应持宽容态度，不应过早介入其他规范，塑造“规则自治→前置法先治→刑法后治”的递进性容错机制。首先，基于生成式人工智能的技术机理，发挥技术规则、行业规则自治对技术缺陷的改进作用；其次，当技术自治失效时，应依次适用民法、行政法化解风险；最后，当滥用生成式人工智能技术行为逾越可容许风险的范围，且重要法益无法得到周延保护时，刑法的介入才具有合法性和正当性。

参考文献：

[1]李颖婷. 生成式人工智能给图书馆带来的机遇、挑战及应对策略[J]. 图书与情报，2023(2)：42-48.

[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(5)：6-13.

[3]生成式人工智能服务管理暂行办法[EB/OL]. [2023-10-10]. https：// www. gov. cn/zhengce/zhengceku/202307/content_6891752. htm.

[4]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.

[5]刘宪权. 人工智能时代的“内忧”“外患”与刑事责任[J]. 东方法学，2018(1)：134-142.

[6]谭影虹. 从数字图书馆到数据图书馆——大数据时代的图书馆服务范式转变[J]. 图书与情报，2016(3)：75-78.

[7]初景利，段美珍. 从智能图书馆到智慧图书馆[J]. 国家图书馆学刊，2019(1)：3-9.

[8]刘月学. 图书馆信息服务生态链构成要素与形成机理研究[J]. 图书馆，2017(6)：53-59.

[9]洪亮，周莉娜，陈珑绮. 大数据驱动的图书馆智慧信息服务体系构建研究[J]. 图书与情报，2018(2)：8-15，23.

[10]邓李君，张晓梅，戴君琴. 图书馆应用人工智能的障碍与对策分析[J]. 图书馆工作与研究，2022(4)：65-69.

[11]ChatGPT plugins[EB/OL]. [2023-10-27]. https：// openai. com/blog/chatgpt-plugins.

[12]王树义，张庆薇. ChatGPT 给科研工作者带来的机遇与挑战[J]. 图书馆论坛，2023(3)：109-118.

[13]Cox C，Tzoc E. ChatGPT：implications for academic libraries[J]. College & Research Libraries News，2023(3)：99.

[14]邓李君，杨文建. 对图书馆应用人工智能的理性思考[J]. 图书馆工作与研究，2021(4)：57-64.

[15]许可. 人工智能的算法黑箱与数据正义[N]. 社会科学报，2018-03-29(6).

[16]丛立先，李泳霖. 生成式 AI 的作品认定与版权归属——以 ChatGPT 的作品应用场景为例[J]. 山东大学学报(哲学社会科学版)，2023(4)：171-181.

[17]深圳市腾讯计算机系统有限公司与上海盈讯科技有限公司著作权权属、侵权纠纷、商业贿赂不正当竞争纠纷案[EB/OL]. [2023-10-27]. https：// wenshu. court. gov. cn/website/wenshu/181107ANFZ0BXSK4/index. html? docId=V4fC-xJ-P6u99sSy9mYo2Pb8FoUruJYuJQluj+GgsVMZxlnEZ3ys8S-J5O3qNaLMqsJiq4jm264lGJFUc7LDfMOufdgNa03jTXfgU-V83pMdA7S0Bq5BLeFV9pq8uck9a6DH.

[18]Terms of use[EB/OL]. [2023-10-27]. https：// openai. com/policies/terms-of-use.

[19]孙山. 人工智能生成内容著作权法保护的困境与出路[J]. 知识产权，2018(11)：60-65.

[20]Lund B D，Wang T. Chatting about ChatGPT：how may AI and GPT impact academia and libraries? [J]. Library Hi Tech News，2023(3)：26-29.

[21]程啸. 民法典编纂视野下的个人信息保护[J]. 中国法学，2019(4)：26-43.

[22]Carlini N，Tramer F，Wallace E，et al. Extracting training data from large language models[EB/OL]. [2023-10-27]. https：// arxiv. org/abs/2012. 07805.

[23]Zhuo T Y，Huang Y，Chen C，et al. Exploring ai ethics of chatgpt：a diagnostic analysis[EB/OL]. [2023-10-27]. https：// doi. org/10. 48550/arXiv. 2301. 12867.

[24]房慧颖. 智能风险刑事治理的体系省思与范式建构[J]. 山东社会科学，2021(2)：187-192.

[25]童云峰. 数字时代图书馆法治化治理模式之提倡[J]. 新世纪图书馆，2022(5)：12-17.

[26]赵蕾，曹建峰. 从“代码即法律”到“法律即代码”——以区块链作为一种互联网监管技术为切入点[J]. 科技与法律，2018(5)：7-18.

[27]马长山. 人工智能的社会风险及其法律规制[J]. 法律科学(西北政法大学学报)，2018(6)：47-55.

[28]童云峰. 个人信息保护法与侵犯公民个人信息罪的衔接机制[J]. 中外法学，2024(2)：366-385.

[29]杨玉晓. 人工智能算法歧视刑法规制路径研究[J]. 法律适用，2023(4)：86-94.

[30]于冲. 刑事合规视野下人工智能的刑法评价进路[J]. 环球法律评论，2019(6)：40-57.

[31]互联网信息服务深度合成管理规定[EB/OL]. [2023-10-25]. http：// www. cac. gov. cn/2022-12/11/c_16722219493-54811. htm.

[32]新一代人工智能发展规划[EB/OL]. [2023-10-27]. http：// www. gov. cn/zhengce/content/2017-07/20/content_52119-96. htm.

[33]童云峰,欧阳本祺.区块链时代智能合约刑事风险的教义学限制[J].西安交通大学学报(社会科学版),2022(2):149-157.
[34]刘宪权.人工智能时代的刑事风险与刑法应对[J].法商研究,2018(1):3-11.
[35]赵秉志,詹奇玮.现实挑战与未来展望:关于人工智能的刑法学思考[J].暨南学报(哲学社会科学版),2019(1):98-110.
[36]雷磊.新科技时代的法学基本范畴:挑战与回应[J].中国法学,2023(1):65-84.
[37]童云峰,欧阳本祺.我国教育法法典化之提倡[J].国家教育行政学院学报,2021(3):26-34,75.
[38]童云峰.大数据时代网络爬虫行为刑法规制限度研究[J].大连理工大学学报(社会科学版),2022(2).

作者简介:
童云峰(1992—),男,特聘副研究员,华东政法大学中国法治战略研究院,上海,201620;
张学彬(1999—),男,西南政法大学法学院2024级刑法学专业在读博士研究生,西南政法大学法学院,重庆,401120。

Legal Fault Tolerance Mechanism of Generative Artificial Intelligence Technology Application in the Field of Smart Libraries

Tong Yunfeng, Zhang Xuebin

Abstract Taking ChatGPT as an example, this article analyzes the technical and legal risks of generative artificial intelligence technology application facing in the smart library field, and proposes the need to build legal fault tolerance mechanism for the application of generative artificial intelligence technology in the library field: the widespread application of self-discipline rules in the first order to control technical risks; the main application of pre-existing laws in the second order to regulate application risks; the progressive application of the third level of criminal law to prioritize minor offenses over serious offenses.

Keywords Smart library; Generative artificial intelligence; ChatGPT; Legal fault tolerance mechanism; Proportionality principle
Class Number G250.76

(上接第22页)

[16]李海俊.数据生产力:主体异化与解放的生产力[J].西北民族大学学报(哲学社会科学版),2021(5):11-19.
[17]戴艳清,戴蒋灿,完颜邓邓.基于云技术的公共数字文化服务协调机制研究[J].情报资料工作,2020(2):93-98.
[18]斯蒂格勒.技术与时间 爱比米修斯的过失[M].裴程,译.南京:译林出版社,2012:30.
[19]谢秋山,陈世香.中西部农村公共服务数字化转型面临的挑战及其应对[J].电子政务,2021(8):80-93.
[20]全国首家乡村智慧图书馆在江夏运营[EB/OL].[2024-04-10].https://news.hubeidaily.net/mobile/z_319548.html.
[21]江维国,胡敏,李立清.数字化技术促进乡村治理体系现代化建设研究[J].电子政务,2021(7):72-79.
[22]李国新.公共文化数字化建设的新方向新任务[J].中国图书馆学报,2022(4):20-22.

作者简介:
王 勇(1991—),男,馆员,福建师范大学数学与统计学院,福建,福州,350117。

Digital Technology Empowering Rural Public Cultural Services: Role Mechanisms, Reality Dilemmas and Optimization Paths

Wang Yong

Abstract The article explains the mechanism of digital technology empowering rural public cultural services from three dimensions: grid digital systems, new public cultural spaces and data precipitation analysis. It points out that the process of digital technology empowering rural public cultural services faces challenges such as weak digital infrastructure, data heterogeneity, data silos, and excessive dependence on technology. The article also proposes corresponding optimization paths, which is to promote the construction of rural digital infrastructure and consolidate the development foundation; build standardized data governance system to enhance the efficiency of digital technology applications; improve the data sharing mechanism and promote the orderly flow of cultural resources; build decision-making mechanism that deeply integrates technological rationality and value rationality, demonstrate humanistic care.

Keywords Rural public cultural services; Digital technology; Service supply
Class Number G240