图书馆虚拟数字人的应用风险、治理困境及责任机制

黄丽
（浙江理工大学法政学院，杭州 311199）

摘要：虚拟数字人技术是元宇宙图书馆落地应用的重要切入点，在为读者带来全新阅读体验的同时也伴随着算法偏见风险、数据安全风险和侵权责任风险，其风险治理面临前期风险难以预测、后期风险难以控制的科林格里奇困境。研究分析图书馆虚拟数字人存在的算法偏见风险、数据安全风险和侵权责任风险，指出技术优化型破解机制存在局限性，提出主体责任型破解机制存在可行性，并以算法偏见风险、数据安全风险、侵权责任风险为治理模块，将图书馆、开发者、读者、监管机构和图书馆虚拟数字人视为行动者，通过行动者网络理论建构图书馆虚拟数字人风险治理的主体责任机制。

关键词：虚拟数字人；生成式人工智能；科林格里奇困境；行动者网络；责任机制
中图分类号：G252；G258.2　DOI：10.3772/j.issn.1673-2286.2024.09.007
引文格式：黄丽. 图书馆虚拟数字人的应用风险、治理困境及责任机制[J]. 数字图书馆论坛，2024，20（9）：64-71.

从技术层面看，虚拟数字人可以理解为通过计算机图形学、图形渲染、动作捕捉、深度学习、语音合成等技术手段创设，并具有人的外观行为甚至思想（价值观）的可交互的虚拟形象[1]。虚拟数字人技术应用已涉及电商、新闻、医疗、教育、文旅等领域，为产业升级、降本增效提供助力。国内图书馆也相继采用了虚拟数字人服务方式，为智能问答、阅读推荐、用户教育等用户服务领域赋能，在一定程度上改变了图书馆的知识生产和服务模式，形成以数据为核心的智能化知识生产与服务生态[2]。

相较于广泛应用的实践现状，虚拟数字人的理论研究明显滞后。图书馆虚拟数字人的研究以技术视角为主，例如：在虚拟数字人的数字记忆概念模型——“记忆数字人”的基础上，讨论元宇宙中“真人数字人”的数字记忆、数字身份认同和价值体系建构[3]；从元宇宙视角切入，讨论虚拟数字人如何赋能图书馆用户服务[4]；还有学者尝试采用GitHub，在QQ上部署虚拟数字人“江小喵”，模拟真人用户为QQ好友或群组提供图书馆相关服务[5]。

在虚拟数字人为图书馆和读者带来全新阅读和服务体验的同时，随之而来的风险不可忽视。如何处理创新进步与安全发展间的冲突，是当下应着力解决的难题。然而，目前我国尚未制定针对虚拟数字人的专门法律规范，而是根据虚拟数字人涉及的技术领域、法律关系进行规范治理，例如针对生成式人工智能、深度合成技术的算法备案制度，针对数据安全领域的数据安全评估制度等。此外，虚拟数字人的人机协同机理与去中心化属性（即在不同用户端根据用户需要生成不同服务内容）使得主体责任无法有效界分，各方主体责任机制亟待厘清，关于图书馆虚拟数字人责任机制、风险治理的研究尚属空白。因此，本文从图书馆虚拟数字人应用风险切入，探讨技术治理的科林格里奇困境及其优化

收稿日期：2024-04-07

方案的局限性，并引入行动者网络理论，尝试构建图书馆虚拟数字人的责任机制，在一定程度上消解新技术带来的风险和挑战。

1　图书馆虚拟数字人的应用风险

图书馆虚拟数字人是一项综合应用，融合了算法推荐、深度合成、情绪计算、虚拟现实等前沿技术，其应用过程主要存在算法偏见风险、数据安全风险和侵权责任风险。

1.1　算法偏见风险

图书馆虚拟数字人的重要服务场景之一是阅读推荐，包括智能推荐阅读书目、自动生成阅读计划、智能解析读书内容等服务，以算法推荐技术（可与生成式人工智能技术融合）为驱动[6]。算法推荐技术可在读者未提出明确需求的情况下，根据读者画像和信息检索行为实时分析用户的信息需求，智能化地向读者推荐匹配资源，提升信息利用效率和用户满意度[7]。图书馆应用算法推荐技术的步骤包括：通过数据获取形成用户画像，根据画像进行知识关联挖掘，借助融合信息完成知识发现[8]。

算法推荐技术应用中算法偏见是不可避免的[9]，包括理解偏见、数据集偏见、技术偏见和实践偏见。理解偏见是因算法设计者对目的、方法、模型的构思设计与预期目标不一致而产生的偏见，例如算法设计者希望获取读者阅读兴趣数据，对读者借阅图书的名称进行采集分析，但未提取作者、图书内容关键词，也未对读者的性别、年龄等进行关联分析，算法模型依此计算出结果的准确性存疑。数据集偏见是指用于对比分析的数据集本身存在偏见，例如把曾经借阅过不婚主义书籍的读者列入不婚主义书单目标推送人群数据集。技术偏见是指受系统、算力、模型等技术限制，或放大前置偏见效果，或过度简化推荐逻辑，例如将多次借阅女权主义书籍的读者性别推定为女性。实践偏见则是指与用户交互过程中产生的偏见，来自算法给人类呈现信息子集的过程[10]。

算法偏见会引致两种后果：①推荐结果的过度偏差，导致虚拟数字人服务的低质化，其真实性、可信任度存疑；②推荐结果的过度精准，使读者形成信息茧房。前者影响图书馆的读者服务体验，而后者则与图书馆的使命和价值相悖。

1.2　数据安全风险

图书馆虚拟数字人的重要服务场景之二是智能问答。由于生成式人工智能技术具有较强的自然语言理解和生成能力，有学者提出通过生成式人工智能赋能图书馆虚拟数字人应用[11]。智能问答服务中数据的作用贯穿于训练、存储、处理和生成的全过程，数据安全是应用基石，其风险体现在3个方面。

（1）数据非法处理风险。图书馆虚拟数字人处理用户个人信息时，根据《中华人民共和国个人信息保护法》规定，应取得个人同意。图书馆虚拟数字人通过传感器、摄像头获取用户的行动路径数据，通过网页埋点技术获取用户的注册信息、借阅信息、阅读时间、阅读习惯等数据，而上述数据已在司法案例中被认定为个人信息，经过用户同意后方能处理。若采集用户的人脸、指纹等敏感数据，还应获得单独同意[12]。此外，虚拟数字人根据用户指令对互联网公开数据进行抓取时，数据来源合法性难以保证。

（2）数据泄露风险。图书馆虚拟数字人通过生成式人工智能提供服务时，其训练模型的数据基本源自馆内的文献资源库、服务场景及用户个人数据，模型中存储的个人信息、文献数据或模型通过上下文语义所挖掘的数据等均存在泄露风险[13]。读者未及时关闭电脑或者退出账号等疏忽行为，图书馆内部人员的数据保护意识不强或者故意出售读者数据牟取利益的行为，以及图书馆数字技术服务商使用数据时造成的多节点暴露[14]等均可能引致泄露风险。

（3）数据内容风险。以生成式人工智能为支撑的虚拟数字人在生成文本时会产生机器幻觉现象，体现为真实与虚假内容混合的低质量数据[15]。图书馆虚拟数字人不仅仅是问答工具，对于未成年人群体而言，其价值观形成和知识结构搭建会受到虚拟数字人的影响。当提供低质量数据时，虚拟数字人对未成年人的知识积累和价值塑造会产生负面影响。

1.3　侵权责任风险

从当前的实践案例来看，虚拟数字人涉及形象（肖像）侵权、声音侵权、名誉侵权等风险。

（1）形象侵权风险。在魔珐公司诉四海公司虚拟数字人案中，杭州中院认为“虚拟数字人Ada的表现形式借鉴了真人的体格形态，同时又通过虚拟美化的手法表达了作者对线条、色彩和具体形象设计独特的美学选择和判断，构成美术作品。使用Ada形象的相关视频分别构成视听作品和录像制品”。米哈游诉伊秀网络案中，广州互联网法院认定“YOYO鹿鸣”这一虚拟形象符合作品独创性的要求，且具有显而易见的可复制性，属于著作权法中具有独创性的美术作品。未经许可使用他人创作的虚拟数字人形象构成著作权侵权。如果虚拟数字人采用真人形象，还可能涉及被拟态的自然人肖像侵权风险。

（2）声音侵权风险。图书馆虚拟数字人的声音来源分为两类：①自然人声音提取，自然人声音经过加工后合成为虚拟数字人的声音，在朗读书籍、播报新闻时有更好的音频效果；②完全由数字技术生成的机器声音。自然人拥有其声音的人格权，声音的财产权可由双方合同约定；数字技术生成的机器声音不具有人格权，根据合同约定采用知识产权或者数据利益保护。未经许可使用自然人声音作为虚拟数字人声音训练样本的或构成人格权侵权；机器声音的直接复制使用可能构成数据财产权侵权。

（3）名誉侵权风险。图书馆虚拟数字人虽然不享有名誉权，但对其形象的破坏会给应用方（图书馆）带来负面的社会影响，虚拟数字人的名誉保护延伸至真实世界主体的名誉利益[16]。此外，虚拟数字人还存在侵害他人利益的情形，例如徐州某医院导诊机器人被破坏事件、人工智能导航失误等。

上述风险给侵权责任认定带来挑战：虚拟数字人能否视为责任主体？开发者、图书馆承担何种责任？基于用户指令造成侵权的如何分配责任归属？对此，欧盟在起草《人工智能法》的同时，还起草了《人工智能责任指令》，制定成员国应用人工智能技术的侵权责任规则，我国图书馆的虚拟数字人责任机制有待建构。

2 图书馆虚拟数字人的风险治理困境

英国技术哲学家大卫·科林格里奇（David Collingridge）曾指出：“一项技术的社会后果不能在技术早期被准确预见，当技术产生不良后果时，它往往已经成为了整个经济和社会结构中难以抽离的一部分，以至于难以对它进行控制。”学界称之为“科林格里奇困境”[17]。图书馆虚拟数字人的治理正如科林格里奇困境一般：其早期风险难以预防，后期风险难以控制。如何在技术应用前期提升风险预测能力，在技术应用后期强化风险控制能力，是图书馆虚拟数字人治理面临的主要问题。

2.1 图书馆虚拟数字人风险治理难点

（1）图书馆虚拟数字人的定位困境导致风险责任主体难以确定。目前关于虚拟数字人定位的观点众说纷纭，可归纳为主体说、客体说和拟制说。主体说认为由于虚拟数字人已经具备类人的思维能力和服务能力，甚至在多领域已经超越人类的能力和属性，应当赋予虚拟数字人法律主体地位，从而解决责任承担问题。客体说认为虚拟数字人不具备主体的人格属性要求，且实际无法承担法律后果，因而无法成为法律主体。拟制说认为虚拟数字人可以成为类似公司法人的法律拟制主体。根据费英格的人格化拟制思想，将虚拟数字人与人进行比较后，如果虚拟数字人有类人属性，则可以被视为人而获得法律主体地位[18]。法律拟制主体较为典型的应用是公司法人，因其财务责任独立化有利于商业发展而被设置。当虚拟数字人的责任承担独立化有利于社会发展时，虚拟数字人就具备了成为拟制主体的前提条件。图书馆虚拟数字人目前尚无具备独立财务资格的需求和必要，仍然需要由相关人类主体辅助承担责任。但虚拟数字人的自学习和自体涌现等特征使其能够脱离人类主体的掌控，在风险产生的自发性和责任承担的限制性之间的冲突问题造成了当下图书馆虚拟数字人的风险治理障碍。

（2）图书馆虚拟数字人技术的不可预测性导致前期风险难以预防。符号主义人工智能凭借以数理逻辑为基础的推理计算，通过公式、算法的事先设定，生成可预测的优化结果；虚拟数字人内嵌的生成式人工智能与之不同，属于连接主义人工智能，采用神经网络的学习方式，通过大量的已标注实例学习，逼近未知的学习目标，该过程具有不可预测性[19]。不可预测性具体表现为：数据处理目的和数据使用范围无法预测，难以满足法律规范对数据处理目的、手段、范围等内容的透明度要求；算法和参数设置根据用户指令进行动态调整，生成内容会部分脱离开发者的控制范围，体现为虚拟

数字人的双向动态算法黑箱，使算法备案制度、安全评估制度等预防性治理手段流于形式。

（3）图书馆虚拟数字人技术的自体涌现性导致后期风险难以控制。学界普遍认为生成式人工智能技术已经达到通用人工智能的起点，以自体涌现性为重要标志，即虚拟数字人能够依托深度学习能力实现开发者未预先设计的功能。自体涌现性强化了虚拟数字人的智能属性，却带来了后期风险难以控制的治理难题。投喂给虚拟数字人大模型训练学习的数据已通过深度学习机制渗透至模型内神经网络的各个节点，不仅影响了学习样本还影响了模型参数，使得算法和参数与初始设计偏离而发生自体涌现。而且大模型涉及的网络节点数量庞大、相互交错耦合，训练数据对下游的各个节点产生了不同程度的影响，无法进行逆转操作，即便训练数据包含侵权内容也无法通过数据删除来消除影响[20]。虚拟数字人的学习特征使得风险在发生后难以控制，且无法通过技术手段恢复原状。

2.2 技术优化型破解机制的局限性

技术优化型破解机制是技术本位的改进方案，以科林格里奇提出的“技术三性”方案为典型，即在应用前期保证技术的可改正性、可控制性与可选择性：可改正性是指虚拟数字人设计中的关键环节未来可以被改正，将来有调整空间；可控制性是指虚拟数字人系统可根据风险后果进行反馈，并可控制风险发生；可选择性是指对虚拟数字人做出的关键决策，治理者拥有最终选择权。设计研发时，如果图书馆虚拟数字人能够满足上述3个条件，那么即便是初期并未采取治理措施，在技术成熟后风险发生时，治理者也可以通过对技术的改正、控制和选择来规避危险的后果。技术优化型破解机制在传统技术治理中有其合理性，但在图书馆虚拟数字人技术治理中存在局限性。

（1）技术优化型破解机制未形成有效闭环。首先，可改正性、可控制性与可选择性对技术开发者提出了更高的要求，但从目前虚拟数字人责任认定规则来看，开发者只承担特定时段内的维护运营责任而无须承担后期风险治理的责任，基于利益最大化的理性人思维，开发者在前期并没有动力来投入更多成本以实现技术的可改正性、可控制性与可选择性。其次，哪些部分属于关键决策、关键环节，哪些风险可以被反馈等问题在虚拟数字人技术背景下有待解决，图书馆虚拟数字人技术的不可预测性和自体涌现性意味着开发者对虚拟数字人风险难以进行技术型预防治理。

（2）技术优化型破解机制忽视了主体的能动价值。技术本位观侧重对技术的理解和改造，然而图书馆虚拟数字人应用的核心是改造主体间关系，技术只是维系多个主体间关系的纽带。主体间关系包括图书馆与读者、图书馆与内部馆员、图书馆与第三方服务商、图书馆与监管机构之间的合作博弈。技术优化型破解机制可以在一定程度上为将来的风险治理留下余地，但能否顺利实现仍然取决于主体的意愿和制度的激励，因此技术优化型破解机制难以单独发挥作用，需要由主体责任型破解机制主导和推动。

2.3 主体责任型破解机制的适用性

主体责任型破解机制是主体本位的优化方案，以布鲁诺·拉图尔（Bruno Latour）等提出的行动者网络理论[21]为典型，强调主体的互动关系在技术治理中的独特价值，在技术发展初期构建各主体的权责关系，利用主体间的博弈关系和趋利避害的主观能动性倒逼各主体采取行动以预防或者避免风险的发生。行动者网络理论有3个核心概念：行动者、异质性网络和转译。行动者是指希望通过改变现状来改变自身状态的人类或非人类；异质性网络是行动者行为发生的场域；转译是为了联结行动者、实现相互理解而输出内容的过程。该理论已经被充分应用于人工智能技术治理、数字平台治理等领域。基于数字技术的相通性，行动者网络理论在图书馆虚拟数字人风险治理中具有适用性。

（1）图书馆虚拟数字人作为网络中的行动者，可破解定位困境。首先，行动者网络中的行动者不仅包括人类主体，还包括非人类主体，图书馆、读者、开发者、监管机构以及虚拟数字人均可作为行动者参与网络构建。其次，虚拟数字人应定位于异质性网络中心，虚拟数字人与图书馆、读者、第三方建立了数据交换关系和数据处理服务关系。从数据流通的角度来看，虚拟数字人是数据交换的中心节点，多方协同建立合纵连横的网络拓扑关系。从数据处理服务角度来看，虚拟数字人为其他主体提供服务，其他主体也为虚拟数字人的服务提供基础条件，多方形成彼此依赖、共生共融的生态系统。行动者通过改变交换和处理关系的内容来改变自身在异质性网络中的地位和价值，并进一步影响网络的架构模式。通过异质性网络来分析主体的责任机制，是以动态

化和网络化的视角分析原本静态单一的主体关系，更符合虚拟数字人技术的特点和发展规律。尽管图书馆虚拟数字人有类人性和超人性：一方面不断缩小与人类思维方式的差距，另一方面有超越人类的数据分析能力和关联计算能力；但由于虚拟数字人无法独立承担责任后果，虚拟数字人参与网络建构但不成为问责对象。

（2）以行动者权利义务为决策指引，破解前期风险难以预测困境。行动者网络机制下的主体权利义务的设定如同行动指南，为行动者的决策行为提供有效指引。国外多家图书馆针对用户编写了隐私保护指南，其中加拿大《不列颠哥伦比亚省公共图书馆隐私保护指南》、澳大利亚《图书馆隐私保护指南》、加拿大《安大略省公共图书馆使用RFID标签指南》、英国图书情报专业人员协会《图书馆用户隐私指南》等比较有代表性[22]。上述隐私保护指南除了告知隐私处理权利、履行透明度义务外，还对读者、第三方使用隐私信息的行为进行了规范，即通过权利义务的事先设定引导相关行动者采取合规措施以规避风险。目前国内公共图书馆还未针对虚拟数字人编写指南，可借鉴隐私保护指南的权利义务规制思路，针对算法偏见风险、数据安全风险和侵权责任风险分别制定责任机制，结合《中华人民共和国民法典》《中华人民共和国数据安全法》《中华人民共和国网络安全法》《中华人民共和国个人信息保护法》《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》《互联网信息服务算法推荐管理规定》《中华人民共和国公共图书馆法》对图书馆虚拟数字人相关行动者的权利义务进行设定，通过事前的行为指引破解前期风险难以预测困境。

（3）通过网络中行动者间的互动转译以及三重风险循环互动，破解后期风险难以控制困境。图书馆虚拟数字人应用的设计初衷是解决与读者沟通效率低下的问题，与生俱来的转译属性使其在行动者网络中集中了海量的数据资源；海量的资源又进一步反哺了虚拟数字人的转译能力，转译能力的动态提升为后期的风险识别奠定基础。算法偏见风险、数据安全风险和侵权责任风险之间相互影响：数据集的质量问题会引发算法的合理性风险，算法的疏忽大意问题会导致数据安全风险，算法和数据风险都会引发侵权责任问题。三重风险虽然无法涵盖图书馆虚拟数字人面临的全部风险，但通过三重风险循环互动的治理责任机制，可进一步化解后期风险难以控制的困境：当行动者间的责任机制明确时，基于趋利避害的能动性，网络中的主体会各司其职，采取行动以规避风险的发生或致力于损害的降低，进而实现风险控制与治理的目标。

3　图书馆虚拟数字人风险治理的主体责任机制

行动者网络理论以技术与社会相互建构的认识观为前提，将所有参与技术建构过程并试图改变网络状态的人或非人类视为行动者。据此理论构建的图书馆虚拟数字人责任机制见图1。对图书馆虚拟数字人应用治理负责的主体不限于开发者、监管机构，还应当包括图书馆、读者以及虚拟数字人。通过扩大参与技术建构的行动者网络，将广泛的行动者纳入治理网络，赋予各主体在技术建构过程中控制和治理技术的责任[23]。在此基础上，可进一步区分图书馆内外部场景：内部场景只涉及图书馆内部工作人员的责任分工，包含资源组织和服务管理两类应用场景；外部场景涉及图书馆外其他主体的责任划分，包含智能问答和阅读推荐两类应用场景。此外，当情感计算技术被大量应用于图书馆虚拟数字人时，虚拟数字人与人的情感和价值逐渐对齐，此时应将伦理治理作为基底提前介入。

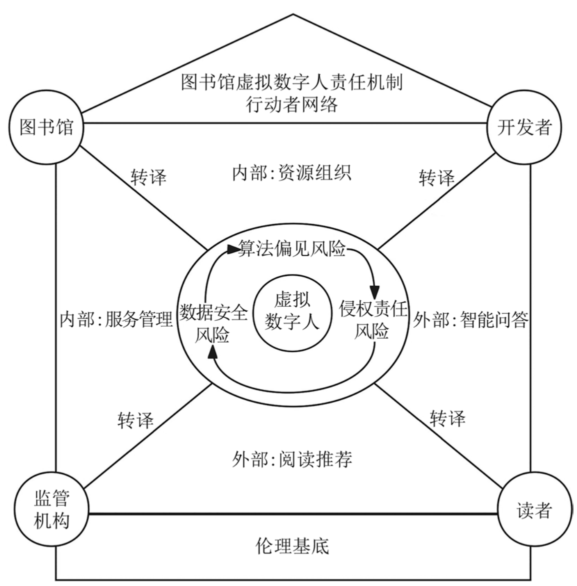

图1　图书馆虚拟数字人责任机制

3.1　算法偏见风险的治理责任

开发者作为算法的设计者应承担主要的治理责

任，具体表现为设定和履行相关义务。①算法透明度义务。个人信息处理者应保证决策的透明度：如须获取读者的个人信息进行算法推荐，应获得读者同意；以显著方式告知读者其提供算法推荐服务的情况，并以适当方式公示算法推荐服务的基本原理、目的意图和主要运行机制等。②算法公正性义务。通过采取措施避免设计、部署和使用环节的缺陷和漏洞，预防算法应用产生的意识形态、社会公平、道德伦理、信息安全等风险。③算法备案义务。部分具有舆论属性或者社会动员能力的算法推荐服务提供者应对算法进行备案。④特殊人群权益保护义务。虚拟数字人服务应当充分考虑老年人、残疾人的需求，避免对老年人、残疾人的日常生活造成障碍；注意设计青少年模式，不推荐不利于未成年人身心健康的书籍，并加入防沉迷机制。⑤算法留痕义务。设计者应依法留存算法网络日志并配合监管，开展评估和监督检查工作。

图书馆提供读者信息和资源信息供开发者进行数据处理和训练，属于个人信息处理者，也应当承担次要治理责任。①有效规避信息茧房。可通过线上协议或者其他便捷方式让用户可自行选择或拒绝使用算法推荐服务，若用户同意使用算法推荐服务，应赋予用户选择兴趣和偏好内容、拒绝其他弱关联书籍推荐的权利。②履行算法透明度义务。可在图书借阅处等位置张贴海报，公示收集用户个人信息的内容、类型，并用简单易懂的方式展示算法推荐的基本原理。③建立健全投诉监督机制。采取措施有效解决读者在使用服务过程中遇到的各种算法问题，并提供通畅的投诉监督渠道。

读者与虚拟数字人的交互会对算法产生影响，读者应承担部分治理责任。①履行目的正当性义务。读者使用图书馆虚拟数字人的目的应正当合法，不得引导虚拟数字人学习或者生成违法信息；不采用机器爬虫等方式批量抓取图书资源、推荐结果等信息，未经许可不得商业化使用虚拟数字人服务。②发现违法违规的虚拟数字人应用行为，应向有关部门投诉、举报。

监管机构在图书馆虚拟数字人应用产业发展中起引导和协调作用，承担相应治理责任。一方面完善算法评估制度，针对图书馆虚拟数字人的公共服务属性和舆论属性进行特殊化的算法评估改造，将涵盖法律规定、技术伦理与开发者承诺的标准视为开发者主观过错的认定标准；协调评估制度与算法透明和算法解释的关系，设置覆盖算法自动化决策全生命周期的问责点[24]。另一方面强化算法备案制度，尽管图书馆虚拟数字人具有舆论属性，但由于公共图书馆多采用同一家企业开发的产品，建议仅要求开发者履行算法备案义务，以减轻图书馆单独备案的行政负担。

3.2　数据安全风险的治理责任

开发者在研发场景、数据收集场景、模型训练场景和技术应用场景下对数据都有一定程度的控制和处理行为，因此承担主要的数据安全风险治理责任。①个人信息保护义务。开发者处理读者个人信息应遵循最小必要原则，加密传输存储、及时脱敏，并完善个人信息保护管理制度。②数据安全保护义务。开发者应不断提升数据加密技术、加强内部工作人员数据安全管理、完善数据流转流程。③数据内容治理义务。开发者应对收集的原始数据、标注数据、训练数据的内容真实性和质量进行评估，尽量选择高质量的数据集，例如：与图书馆合作获得正版图书、高水平论文数据进行训练，减少数据源污染；对生成内容建立初步审核机制，避免生成有害内容；履行内容标识义务，以显著、醒目的方式对合成内容进行标识。④数据来源合法义务。开发者应核查通过交易获得的数据来源，对通过机器抓取的数据应进行合法性分析，避免使用未经授权的个人信息、商业秘密或者含有隐私内容的数据。⑤数据监督义务。委托或共同处理数据时，应对合作者的数据处理行为进行监督。

图书馆作为信息资源的提供方和读者信息的收集方，是原始数据初始的控制者和处理者，对数据安全风险治理承担主要责任。①个人信息保护义务。在收集读者个人信息时应获得读者同意；在有摄像头以及采用人脸识别设备采集敏感数据时应获得单独同意；传输个人信息时应当加密，并遵循最小必要原则。②数据安全保护义务。应对读者和馆藏图书资源数据进行加密保护，并尽量保存在本地，如果必须存储在他处应考虑灾备问题；对图书馆工作人员进行培训，强化数据安全保护意识。③数据监督义务。在委托或共同处理数据时，应对合作者的数据处理行为进行监督。

读者在使用图书馆虚拟数字人服务时，会上传个人信息或提供部分数据，也应承担数据安全风险治理的部分责任。①个人信息管理义务。读者应在安全稳定的环境中使用服务，不下载有监视、爬虫功能的终端应用，及时清除浏览器输入信息记录，妥善保管使用记录，避免外泄。②数据保护义务。不上传他人数据或商

业秘密数据至虚拟数字人对话框、不大量抓取虚拟数字人输出的馆藏资料、未经许可不将生成内容用于商业目的、不随意将虚拟数字人提供的数据传输至境外。③应用保护义务。不使用外挂、模拟账户等访问虚拟数字人服务器，保证虚拟数字人平稳正常提供服务。如果发现虚拟数字人的数据泄露问题，读者应及时向图书馆或者相关部门反映。

3.3　侵权责任风险的治理责任

开发者在研发设计虚拟数字人时，会涉及知识产权、声音、肖像侵权等问题，对此承担首要的治理责任。①美术作品侵权排查。开发者在设计虚拟数字人形象时，应排除与已有美术作品形象的实质性近似，尤其是知名的卡通漫画形象、商标图案等。②肖像侵权排查。开发者使用真人图片加工处理超写实虚拟数字人时要做近似比对，如果产生与某自然人的关联情况，仍应获得授权。③声音侵权排查。用真人声音作为虚拟数字人音源的应当获得本人授权同意。④数据集知识产权侵权排查。开发者使用图书进行训练时，应当获得相应知识产权许可。

图书馆作为部署并实际使用虚拟数字人服务的主体，有能力对开发者形成制约，但考虑到图书馆的公共属性，不应对图书馆赋予过重的主体责任，影响图书馆提升服务质量的积极性。建议在侵权责任风险治理时对图书馆采用避风港原则，图书馆承担以下治理责任。①“通知—删除”义务。当权利人通知图书馆虚拟数字人服务侵权并提交初步证据时，图书馆在确认证据的有效性后应尽快停止相关功能，以避免侵权损失或影响扩大。②披露协助义务。在图书馆接到权利人侵权通知后，应当及时披露开发者信息供权利人起诉维权，并帮助权利人完成相关取证操作。③监督审核义务。图书馆采购服务时应严格审查开发商资质和过往案例，服务验收时审查知识产权合法性证明材料。④知识产权授权义务。图书馆应先取得图书等资料的知识产权许可，再进行大模型训练，以免造成知识产权侵权，如果虚拟数字人服务只涉及图书简介内容，可基于合理使用阻却侵权风险。

以目前国内公共图书馆应用较为广泛的“博看数字人”为例，开发者是博看公司，应用者是各地公共图书馆，二者在应用部署前应针对虚拟数字人相关的责任内容签署协议，主要条款为：博看公司承担算法偏见风险、数据安全风险和侵权责任风险的主要治理责任，履行相关义务；图书馆在算法偏见风险治理中承担次要责任，在数据安全风险治理中承担主要责任，在侵权责任风险治理中采用避风港原则。此外，各地公共图书馆在上线虚拟数字人服务时应同时上线一份读者协议，明确读者个人信息处理的目的、过程等信息，履行透明度义务；同时要求读者秉持合法良善目的使用服务，并保护相关个人数据以免发生泄露等。

4　结语

图书馆虚拟数字人能够提升图书馆的服务水平，为读者带来绝佳的阅读体验，其带来的风险问题亦不容忽视。本文以图书馆虚拟数字人应用面临的算法偏见风险、数据安全风险、侵权责任风险为研究对象，以解决应用风险治理的科林格里奇困境为研究目标，对破解困境的技术优化型方案和主体责任型方案进行论证和比对，最终选择了行动者网络理论下的主体责任型方案。以行动者网络理论为基础构建的图书馆虚拟数字人责任机制立足于开发者、图书馆、读者和监管机构的视角，并将虚拟数字人视为行动者之一，不断明晰行动者网络中主体的价值目标和责任内涵，采取最优行动策略以完成目标。责任机制还可以用于标准制定、合同签署、政策出台，以及图书馆相关指南制定。

参考文献

[1]　张丽锦，吕欣．虚拟数字人：模因论的新“锚点”：模因论视域下的虚拟数字人：概念、特征和应用[J]. 学术探索，2024（3）：57-66.

[2]　李冬梅．图书馆虚拟数字人：内涵特征、信息模型与应用场景[J]. 新世纪图书馆，2023（7）：51-57，73.

[3]　黄薇，夏翠娟，铁钟．元宇宙中的数字记忆：“虚拟数字人”的数字记忆价值[J]. 图书馆论坛，2023，43（7）：154-160.

[4]　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]. 图书馆建设，2023（6）：62-68.

[5]　刘琼，刘桂锋，王鹏．AIGC赋能图书馆阅读推广智慧服务的框架和应用研究[J]. 图书馆学研究，2024（2）：108-118，107.

[6]　徐芳．智慧图书馆生成式人工智能应用场景及其法律问题[J]. 情报资料工作，2024，45（2）：24-29.

[7]　刘海鸥，陈晶，孙晶晶，等．面向大数据的移动数字图书馆情境

化推荐系统研究[J]. 图书馆工作与研究，2018（9）：58-64.

[8] 刘海鸥，李凯，姜波. 移动图书馆推荐系统中的用户画像与资源画像情境化融合研究[J]. 图书馆，2021（6）：66-71，93.

[9] 孟令宇. 从算法偏见到算法歧视：算法歧视的责任问题探究[J]. 东北大学学报（社会科学版），2022，24（1）：1-9.

[10] 贾诗威，闫慧. 算法偏见概念、哲理基础与后果的系统回顾[J]. 中国图书馆学报，2022，48（6）：57-76.

[11] 郭亚军，郭一若，冯思倩，等. ChatGPT赋能高校图书馆元宇宙空间服务[J/OL]. 图书馆论坛：1-10[2024-02-04]. http://kns.cnki.net/kcms/detail/44.1306.g2.20240123.1149.004.html.

[12] 黄丽. 论个人信息单独同意的利益衡量模型（英文）[J]. 科技与法律（中英文），2022（5）：138-148.

[13] 李佳轩，储节旺，杜秀秀. 关联、黑箱与赋能：AIGC驱动智慧图书馆的转型路径[J]. 图书情报工作，2023，67（23）：18-27.

[14] 孙丽敏. 图书馆读者数据泄露及其防范机制摭谈[J]. 河南图书馆学刊，2023，43（8）：120-122.

[15] 漆晨航. 生成式人工智能的虚假信息风险特征及其治理路径[J]. 情报理论与实践，2024，47（3）：112-120.

[16] 颜卉. 算法驱动型虚拟数字人涉侵权纠纷的规范解决路径[J]. 重庆大学学报（社会科学版），2024，30（2）：182-194.

[17] 文成伟，汪姿君. 预知性技术伦理消解AI科林格里奇困境的路径分析[J]. 自然辩证法通讯，2021，43（4）：9-15.

[18] 俎璐. 人工智能法律拟制主体地位再探：面向拟制哲学视角下的法律主体制度[J]. 北京理工大学学报（社会科学版），2024，26（4）：152-163.

[19] 寿步. 论人工智能生成内容的可版权性和版权人问题[J]. 科技与法律（中英文），2024（4）：60-72.

[20] 黄丽. 生成式人工智能训练数据的软硬法协同治理研究[J]. 宁夏大学学报（社会科学版），2024，46（1）：112-121，136.

[21] 吴莹，卢雨霞，陈家建，等. 跟随行动者重组社会：读拉图尔的《重组社会：行动者网络理论》[J]. 社会学研究，2008（2）：218-234.

[22] 田淑娴，许春漫. 国外图书馆用户隐私保护指南文本分析与启示[J]. 图书情报工作，2015，59（18）：61-66，116.

[23] 张欣. 生成式人工智能的数据风险与治理路径[J]. 法律科学（西北政法大学学报），2023，41（5）：42-54.

[24] 张凌寒. 算法评估制度如何在平台问责中发挥作用[J]. 上海政法学院学报（法治论丛），2021，36（3）：45-57.

作者简介

黄丽，女，博士，讲师，研究方向：知识产权法、数据法，E-mail：leavy.hl@zjtu.edu.cn。

Application Risks, Governance Dilemma, and Responsibility Mechanisms of Virtual Digital Human in Libraries

HUANG Li
(School of Law and Politic, Zhejiang Sci-Tech University, Hangzhou 311199, P. R. China)

Abstract: Virtual digital human technology is an important entry point for the implementation and application of metaverse libraries. While providing readers with a brand-new reading experience, it is also accompanied by algorithmic bias risks, data security risks, and infringement liability risks. Its risk management faces the Collingridge's Dilemma, where early risks are difficult to predict and later risks are difficult to control. This article analyzes the algorithmic bias risks, data security risks, and infringement liability risks of virtual digital human technology in libraries, points out the limitations of technology optimized cracking mechanisms, and proposes the feasibility of subject responsibility based cracking mechanisms. With algorithm bias risks, data security risks, and infringement liability risks taking as governance modules, libraries, developers, readers, regulatory agencies, and library virtual digital human technology are regarded as actors, and the main responsibility mechanism for library virtual digital human risk governance is constructed through actor network theory.

Keywords: Virtual Digital Human; Generative Artificial Intelligence; Collingridge's Dilemma; Actor Network; Responsibility Mechanism

（责任编辑：王玮）