

新时代下虚假信息检测与智慧图书馆联合发展模式

张宏彬^{*} 王新雨 石宸

南京理工大学, 中国·江苏 南京 210094

摘要: 如今, 人工智能、深度伪造、社交机器人等技术的不断发展, 虚假信息的生产与传播也愈演愈烈, 已经逐渐对社会知识安全构成严峻挑战。智慧图书馆对虚假信息检测技术的应用, 可以保障图书馆知识安全, 也能反过来为虚假信息检测提供更加良好的知识背景。本文在系统梳理虚假信息检测技术演进脉络的基础上, 阐释其与智慧图书馆赋能与反哺的内在逻辑, 进而提出“双向赋能—循环增强”框架, 最终从制度、技术、信息安全三个维度提出联合推进策略, 为构建未来信息安全式智慧图书馆提供可行路径。

关键词: 虚假信息检测; 智慧图书馆; 知识图谱; AIGC 治理; 可信知识生态

The Joint Development Model of False Information Detection and Smart Libraries in the New Era

Zhang Hongbin^{*}, Wang Xinyu, Shi Chen

Nanjing University of Science and Technology, China Jiangsu Nanjing 210094

Abstract: Nowadays, with the continuous development of technologies such as artificial intelligence, deep fakes, and social robots, the production and dissemination of false information have become increasingly rampant, gradually posing a severe challenge to social knowledge security. The application of false information detection technology in smart libraries can ensure the knowledge security of libraries and, conversely, provide a better knowledge background for false information detection. Based on a systematic review of the evolution of false information detection technology, this paper explains the internal logic of its empowerment and feedback with smart libraries, and then proposes a "bidirectional empowerment - circular enhancement" framework. Finally, it puts forward joint promotion strategies from three dimensions: system, technology, and information security, providing a feasible path for building a trustworthy knowledge ecosystem in smart libraries.

Keywords: False information detection; Smart libraries; Knowledge graph; AIGC governance; Trustworthy knowledge ecosystem

0 引言

互联网进入“后真相”时代, 新型内容生产方式不断发展^[1], 虚假文本、合成图片、深度伪造视频及算法生成音频的产业化生产, 使得“真伪难辨”变得司空见惯。《2024 全球风险报告》指出, 如今生活成本危机持续加剧, 在 AI 赋能^[2]的虚假信息与社会信息安全相互作用下, 社会两极分化加剧, 其中教育、医疗与公共政策等领域问题尤为严重。图书馆作为全民学习的重要机构, 信息真实性的重要性不言而喻, 而其中的数字馆藏、参考咨询、阅读推广等各项服务却都会受到“真伪难辨”的影响。因此, 智慧图书馆需要利用虚假信息检测对纳入馆藏的内容进行严谨的验证, 也需要充当虚假信息检测大规模、高质量、语义关联的权威数据来源, 成为新时代“防谣治谣”的新基建。

1 赋能逻辑

1.1 虚假信息检测技术的功能及其发展

所谓虚假信息检测技术, 是利用大模型、大数据分析、自然语言处理等相关方法发现虚假、误导、恶意信息的一套技术体系, 它主要包括假新闻、社交媒体谣言、深度伪造内容等多种应用场景。

但随着人工智能技术的不断成熟, ChatGPT、豆包等大模型的图像生成技术以及编造虚假信息的能力愈发强大, 而现有的虚假信息识别技术则已经略显乏力。在此背景下, 研究人员不得不进一步针对 AI 生成和 deepfake 等造假技术研发更加有效检测手段。同时, 随着各方平台的虚假信息逐渐泛滥, 社交媒体、事实核查机构以及政府部门都在不断地尝试, 试图将数据和资源以加密的形式共享, 从而构建一个体系化的虚假信息治理网络。

1.2 虚假信息检测技术赋能智慧图书馆的内在逻辑

虚假信息检测技术赋能智慧图书馆内在逻辑是构造一整套自我检验的知识治理体系，把原先针对社交媒体谣言所形成检测能力注入到图书馆数字化生态中作为内生免疫系统的一部分。因为在数字化生成链路过程中，存在一定的技术问题或者人为因素导致文本或者信息出现了错误，从而导致数字化生产的链条末端诞生了大量虚假信息。

从更深层面上来讲，以虚假信息检测为代表的新一代信息技术对智慧图书馆的赋能，本质上就是一场围绕着知识基础设施而开展的自我革命。当图书馆能够利用虚假信息检测技术来审视和检验自身的数字化的时候，图书馆最根本的伦理就已经开始重塑——对知识真实性的终极负责。在此过程中，智慧图书馆不再只是一个简单的信息存储机构，而是为人类知识建立了一个知识服务生态的信任根基^[3]。

2 智慧图书馆在虚假信息检测中的反哺机制

智慧图书馆通过技术为知识赋能，对馆藏资源进行深度治理，不仅能够提升了自身数据的可信度，还能建立一套可靠的知识验真标准，进而能够促进社会信息可信生态的发展（见图 1）。



图1 智慧图书馆反哺虚假信息检测技术逻辑

2.1 知识背景的强化与验证基准构建

在该框架下，智慧图书馆则能通过确保知识的可信性，形成权威的知识背景，在遇到未知的信息时，就可以快速提取图书馆内部知识图谱中的权威信息，快速验证信息真伪。这一个改变，使得图书馆不再仅仅只是一个信息存储的机构，而逐渐附加了社会信息生态中的重要“事实锚点”效果。

2.2 跨域知识图谱的协同验证

通过开放数据接口的方式，智慧图书馆能够与新闻网站、社交媒体等平台共享数据，实现跨领域的虚假信息检测。在该集成模式下，如果发现了一些难辨真伪的信息，虚假信息检测技术则可以立马借助图书馆的知识背景和其他平台的实时信息，对信息进行实体间的逻辑一致性检验

（如时间冲突、内容矛盾）来判别信息真伪。这种跨领域的合作方式，不仅能够降低了单一数据源可能导致的误判风险，也能够扩大了虚假信息检测技术的适用范围。

2.3 社会信息治理的认知干预

除了技术层面的反哺，智慧图书馆还能通过可靠的知识传播帮助公众间接塑造社会的信息认知模式。

2.3.1 可信摘要嵌入

通过图书馆的知识库，图书馆通过社交媒体向读者阅览的信息进行结构化总结，为读者直接提供参考。

2.3.2 争议案例教学

图书馆虚假信息识别中的一些典型虚假信息案例加工成为信息素养教育素材，并用多模态的方式把其分析过程可视化展现给公众。

2.3.3 溯源习惯培养

强化检索服务中出处标记，将用户引向信源背后的验证链条，帮助大众培养主动排查信息真伪的习惯^[4]，形成“权威知识输出一公众意识提升—信息环境净化”的良性循环。

3 构建虚假信息检测与智慧图书馆联合体系框架

虚假信息检测与智慧图书馆的联合体系框架，从基础设施层、数据资源层到业务应用层^[5]，结合了机器学习、社会网络分析、知识图谱等技术以及新闻、社交媒体的数据资源平台，构建了完整的联合体系框架（见图 2），为实现图书馆在权威信息验证、图书馆知识图谱构建和大众信息认知干预方面提供了新的思路。

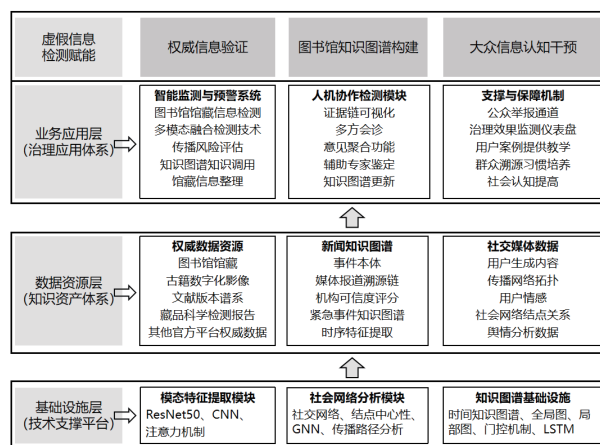


图2 虚假信息检测与智慧图书馆联合体系框架

3.1 基础设施层：构建智能化的技术支撑平台

基础设施层主要包括模态特征提取、社会网络分析和知识图谱基础设施三大核心模块，是整个联合体系框架的技术支撑。其中，模态特征提取模块包括检测相关算法，

为虚假信息检测提供了坚实的技术支撑。而社会网络分析模块实时追踪微博、微信等社交平台的信息传播路径,监测异常信息的传播情况。在知识图谱基础设施方面,该模块将时序知识图谱、全局图和局部图三图融合,再结合门控机制和长短时记忆网络,构建了一套快速结合已有知识背景识别信息真伪的可行性方案。

3.2 数据资源层: 构建多维度的知识资产体系

数据资源层主要在于形成多维化的、高质量的知识资产体系。该模式收集图书馆古籍数字化的影像、文献版本谱系、藏品科学检测报告等内容,经过严谨的质量把控和标准制定,建成可靠的知识基准。在数据治理方面,该模式实施分级分类策略,引入数据安全机制,核心馆藏数据将采用严格的访问控制,社交媒体数据通过差分隐私处理,新闻数据则建立完整的溯源体系,分层次分类别治理这些数据。

3.3 业务应用层: 打造智能化的治理应用体系

业务应用层负责把技术能力转化为治理效能,建立了三个主要的应用系统。智能监测预警系统通过分布式爬虫获取社交媒体热点话题,在多模态融合的基础上,对其进行联合分析,通过知识图谱快速对可疑信息进行验证。人机协作检测模块是基于人机协作设计理念研发的机制,为专家团队提供证据链可视化、多方会诊等功能,能够更好地发挥出人类专家和机器智能的互补性价值。支撑与保障机制则根据虚假信息的性质和传播态势,智能匹配最合适的应对策略,包括在社交平台嵌入权威信息提示、开展定向的信息素养教育等。

该联合框架不仅提高了图书馆本身的信息治理水平,也把图书馆自身的专业优势发挥到了社会信息生态治理当中,把图书馆从原先的知识存储机构转变成了知识治理枢纽,实现了战略定位的升级换代。

4 虚假信息检测联合智慧图书馆的推进政策

在建立虚假信息检测与智慧图书馆联合体系框架的基础上,社会需要从制度策略、技术策略和信息安全策略三个维度上制定系统的推进策略,从而确保该体系的可持续发展。

4.1 制度策略: 构建协同治理长效机制

制度的建立,是推动该联合框架体系长远发展的根本措施。合理的制度,需要由多方共同参与、共同建立,并且不可一蹴而就,是一个不断探索和完善的过程,要严谨在前,而后勇于探索。

4.1.1 建立跨部门联席工作机制

组建由文化和旅游部门牵头、会同网信办、教育局、

科技局等部门成立专项工作组,制定《智慧图书馆虚假信息防控指南》,明确各方职责分工。在保障数据安全的前提下,图书馆还可以与微博、新华社等多家媒体及新媒体矩阵形成战略合作关系,建立信息共享、联合研判机制。

4.1.2 完善行业标准规范体系

在建立和完善相关技术和工作机制的基础上,加快出台《图书馆数字资源真实性认证标准》等标准,将虚假信息检测纳入图书馆评估指标体系,在公共图书馆的有关法律修改时增加数字资源真实性的保障条款等。

4.1.3 设立专项扶持基金

地方政府可根据实际情况设立相应专项扶持基金,国家社科基金、国家自然科学基金可以设立“数字时代知识真实性保障”等专项课题,支持相关关键技术的研究开发及示范工作。

4.2 技术策略: 吸收先进技术, 升级基础设施

在虚假信息检测与智慧图书馆联合体系的构建中,技术策略是实现高效治理的关键。通过引进一些先进的技术手段,图书馆自身的一些设施条件需要做出改善,从而有效提高平台的数据检测水平、数据处理水平和数据的资源利用率。

4.2.1 形成“国家—区域—馆所”三级技术服务平台

国家级平台做好核心算法的研发工作,区域中心做好算力支持工作,各馆所端部署轻量化检测终端,利用好“东数西算”工程算力资源,从而降低技术的应用门槛。

4.2.2 建立分级开放共享机制

将图书馆数据分为核心馆藏(严格保护)、基础资源(有条件共享)、公共服务数据(完全开放)三级,制定差异化开放策略,同时建立好信息安全机制,防止各个机构信息泄露。

4.2.3 打造全国文献真实性数据库

汇集中央各馆古籍版本记录、藏品检测报告等权威性的数据资料,以分布式知识验证网络为基础,运用区块链保障数据有源可循、有据可查、不可篡改^[6]。

4.3 信息安全策略: 筑牢数据隐私保护防线

在引入数据共享的背景下,只有保障了信息安全,数据与资源的共享才能有意义,也才能更好地维护信息生态安全。

4.3.1 实施隐私增强技术

对读者查询记录、社交数据等敏感信息采用联邦学习、差分隐私等技术处理,实现“数据可用不可见”。

4.3.2 健全安全审计制度

建立数字化全流程留痕机制,关键操作需多重认证,

定期开展系统渗透测试和安全风险评估。

4.3.3 培养复合型人才

开设“图书馆信息安全官”岗位，与高校联合培养熟悉图书馆业务，懂得信息安全技术的复合型人才^[7]，开展全员每年至少一次的安全意识培训。

上述策略需要动员政府、行业与馆所等多主体共同努力，在省级或高校图书馆层面先行先试，探索总结经验之后，逐步在全国馆范围内部署推广，进而实现整体上技术先进、制度完善、安全可靠的虚假信息协同治理体系目标，从而为数字中国建设提供高质量的知识基础设施保障。

5 结语

在如今数字化不断发展和人工智能技术不断加强的时代下，虚假信息不断渗透进公共文化机构，虚假信息检测技术的发展也愈发困难。通过智慧图书馆与虚假信息检测技术的联合体系框架，不仅能够保障图书馆数字化过程中的信息安全，还能进一步促进虚假信息技术的发展。本研究不仅研究构思了“技术赋能－知识反哺”的双向循环的可行性道路，还重新确立了智慧图书馆在新时代下的信息生态安全中的地位。关于推进政策方面，各方只有不断完善图书馆知识图谱的更新机制、构建多方参与的协同治理网络、营造良好的信息安全生态网络以及培养大众的信息溯源习惯，图书馆才能真正地成为“知识免疫中枢”，为社会提供一个更加健康、安全的知识生态环境。

参考文献：

[1] 孙榕，李白杨. 生成式人工智能 (GAI) 背景下

的新型数字鸿沟识别框架研究[J]. 图书情报知识, 2025 (3):19-30.

[2] 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):34-43,135-136.

[3] 冀晓萌. 基于知识服务的智慧图书馆知识资源建设研究与实践 [J]. 图书馆学刊, 2025,47(5):45-48,60.

[4] 钱彦，梅影. 从理念到实践：生成式人工智能在智慧图书馆中的应用探索[J]. 图书馆研究与工作, 2023 (12):27-34.

[5] 常强，赵伟. 人工智能生成内容 (AIGC) 技术赋能智慧图书馆发展：逻辑、框架与策略 [J]. 图书馆学刊, 2025,47(6):109-113.

[6] 初景利，任娇菡，王译晗. 从数字图书馆到智慧图书馆 [J]. 大学图书馆学报, 2022,40(2):52-58.

[7] 冯耕. 基于图书馆学专业教育的公共图书馆人才培养研究 [J]. 图书馆学刊, 2025,47(4):42-45.

基金项目：南京理工大学本科生科研训练计划立项资助，项目编号：202410288107Z。

作者简介：张宏彬（2003-），男，汉族，福建莆田人，南京理工大学信息管理与信息系统专业 22 级在读本科生，研究方向：图书馆学、情报分析、舆情分析、知识图谱、多模态数据处理。