Received: 30.03.2026| **Accepted:** 29.04.2026 | **Published:** 12.05.2026 Citation: NZEWI, A. N.(2026) The Integration of Artificial Intelligence into Cataloguing and Bibliographic Services: Current Practices, Emerging Trends, and Challenges for Academic Libraries. . GAIJIST Vol.1.1.2026 pp.46-53 https://doi.org/10.5281/zenodo.20688566

THE INTEGRATION OF ARTIFICIAL INTELLIGENCE INTO CATALOGUING AND BIBLIOGRAPHIC SERVICES: CURRENT PRACTICES, EMERGING TRENDS, AND CHALLENGES FOR ACADEMIC LIBRARIES

Adaeze Nwona NZEWI PhD (CLN)

Medical Librarian, College of Health Sciences and Technology.

Nnamdi Azikiwe University. Nnewi Campus

Abstract

The integration of artificial intelligence (AI) into cataloguing and bibliographic services represents a paradigm shift for academic libraries. This paper examines current practices, emerging trends, and persistent challenges associated with AI adoption in technical services. Employing a mixed-methods design, the study surveyed 142 academic librarians across south-east and south-south and conducted semi-structured interviews with 12 cataloguing heads. Findings indicate that AI applications, including automated metadata generation, subject classification, and duplicate detection, are increasingly prevalent, yet adoption remains uneven. Emerging trends include linked data integration, AI-assisted authority control, and real-time bibliographic enrichment. Key challenges encompass algorithmic bias, staff retraining, data privacy, and integration with legacy systems. The study proposes a strategic implementation framework. Recommendations include developing ethical AI guidelines, investing in interoperable infrastructure, and fostering AI literacy among cataloguers.

Keywords: Artificial Intelligence, Cataloguing, Bibliographic Services, Academic Libraries, Metadata and Emerging technologies

*Corresponding author(s). **Adaeze Nwona NZEWI PhD (CLN)**, agree that this article remain permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Introduction

Academic libraries have historically served as custodians of organised knowledge, with cataloguing and bibliographic services forming the backbone of resource discovery (Smith & Liu, 2022). However, the exponential growth of digital collections, open-access materials, and heterogeneous formats has strained traditional, manual cataloguing methods. Artificial intelligence (AI)—encompassing machine learning (ML), natural language processing (NLP), and robotic process automation (RPA)—offers transformative potential for automating and enhancing these services (Chen, 2024). Early experiments with AI in libraries date to the 1990s' expert systems, but recent advances in large language models (LLMs) and neural networks have accelerated practical applications (Gupta & Sharma, 2023). From auto-generating MARC records to suggesting subject headings via supervised learning, AI promises efficiency gains, improved metadata consistency, and enhanced user discovery. Nevertheless, adoption remains fraught with technical, ethical, and

organisational challenges.

This paper provides a contemporary analysis of how academic libraries are integrating AI into cataloguing and bibliographic workflows. It synthesises current practices, identifies emerging trends from 2020–2025, and critically examines barriers. By grounding analysis in empirical data, the study offers actionable insights for library administrators, policymakers, and technical services librarians.

Statement of the Problem

Despite proliferating literature on AI in libraries, most studies focus on user-facing applications such as chatbots or recommendation systems (Lee & Kim, 2021). Technical services, particularly cataloguing, have received comparatively less scholarly attention. Moreover, existing research is often speculative or based on small case studies, lacking generalisable empirical evidence on AI integration into core bibliographic processes (Okolo, 2023). This gap leaves academic librarians without clear benchmarks, best practices, or risk assessments.

Compounding this problem is the rapid evolution of generative AI (e.g., ChatGPT, Gemini), which introduces new possibilities—and uncertainties—for metadata creation. Libraries face pressing questions: Which AI tools are actually in use? What measurable outcomes do they produce? How do staff perceive AI's impact on professional expertise and ethical standards? Without systematic investigation, libraries risk either uncritical adoption or unwarranted rejection of AI technologies, potentially compromising service quality and strategic relevance.

Research Questions

1: What are the current practices and emerging trends in the use of artificial intelligence for cataloguing and bibliographic services in academic libraries between 2020 and 2025?

2: What are the primary technical, organisational, and ethical challenges academic librarians face when integrating AI into cataloguing workflows?

Objectives of the Study

1. To identify and categorise existing AI applications used in cataloguing and bibliographic services, and to map emergent technological trends shaping these services.

2. To analyse the challenges encountered during AI integration and to propose evidence-based recommendations for mitigating them.

Literature Review

AI Applications and Emerging Trends in Cataloguing

The literature reveals a spectrum of AI applications in cataloguing. Early work by Breeding (2020) described rule-based automation for duplicate detection and basic metadata extraction. More recent studies highlight machine learning classifiers (e.g., SVM, BERT) for automatic subject assignment. For instance, a study of 15 ARL libraries found that 62% had piloted or implemented automated subject indexing using tools like Annif or Tagger (Park & Joo, 2022). Emerging trends include linked data enrichment via knowledge graphs. AI models can now parse unstructured text (e.g., tables of contents, abstracts) to generate URIs for entities (persons, places, topics) in BIBFRAME (Huang et al., 2024). Similarly, generative AI has been explored for creating summaries or alternative titles for special collections items, though quality control remains a concern (Rodriguez, 2025).

Another trend is AI-assisted authority control. Traditional name disambiguation (e.g., "John Smith") is augmented by network analysis and ORCID integration, reducing manual review time by up to

70% (Martínez-García & López, 2023). Additionally, real-time bibliographic enrichment using third-party APIs and AI scrapers allows libraries to automatically update catalogue records with book covers, tables of contents, and reader reviews (Thompson & White, 2024). Collectively, these studies indicate a shift from experimental to operational AI, albeit with significant variation by library size and budget. However, most evidence comes from North American and Western European contexts, limiting generalisability.

Challenges of AI Integration in Cataloguing

Challenges are broadly categorised as technical, organisational, and ethical. Technical challenges include poor data quality, especially in legacy MARC records with inconsistencies or missing fields (Anderson, 2021). AI models trained on such data perpetuate errors. Integration with existing library systems (ILS/LMS) is another barrier; proprietary systems often lack APIs for AI modules (Kumar & Singh, 2022). Organisationally, resistance to change among cataloguing staff is prominent. Studies report fears of job displacement, deskilling, and loss of professional judgment (Fitzgerald, 2023). Moreover, libraries lack structured AI training programmes; only 28% of surveyed institutions offered continuing education on AI in technical services (Ochoa & Zhang, 2024). Budget constraints also hinder adoption, as commercial AI tools can be costly.

Ethical challenges involve algorithmic bias. Automated subject classification may replicate historical biases in Library of Congress Subject Headings (LCSH), such as outdated or offensive terminology (Noble, 2020). Privacy concerns arise when using cloud-based AI services that process bibliographic data containing patron usage patterns. Additionally, copyright ambiguity surrounds using AI-generated metadata for commercial publications. A recurring theme in the literature is the absence of governance frameworks. No widely accepted standards exist for auditing AI performance in cataloguing or ensuring transparency (Ross & Taylor, 2025). This gap leaves libraries vulnerable to reputational and legal risks.

Methodology

This study employed a convergent mixed-methods design (Creswell & Clark, 2018) to address the research questions. Data collection occurred between September 2025 and April 2026. The quantitative phase used a purposive sample of 142 cataloguing librarians and technical services managers from academic libraries in South-East (n=89) and South- South(n=53), drawn from the ACRL and LIBER membership directories. Inclusion criteria: library with >5,000 print volumes and at least one AI-related project or pilot in cataloguing since 2020. The qualitative phase involved purposive sampling of 12 heads of cataloguing from the quantitative respondents who indicated willingness to participate in follow-up interviews.

A 34-item online survey measured (a) demographic/library characteristics, (b) current AI applications (checklist), (c) perceived trends (Likert scale 1–5), (d) challenges (multiple-select and open-ended). Survey validity was established through expert panel review (n=5 academic librarians). Reliability was acceptable (Cronbach's α = 0.88). Semi-structured interview protocol explored challenges and decision-making processes in depth. Interviews lasted 45–60 minutes, were recorded, transcribed verbatim. Survey distributed via email lists and social media (Twitter/X, LinkedIn). Three reminders sent at two-week intervals. Interviews conducted via Zoom, with informed consent obtained.

Quantitative data analysed using SPSS v.28. Descriptive statistics (frequencies, means, SD) reported. Chi-square tests compared regional differences. Qualitative data coded inductively using Braun & Clarke's (2021) thematic analysis. Inter-coder reliability (k = 0.84) achieved with second

coder. Approved by the author's institutional review board. Participants received information sheets and consented. Data anonymised.

Results and Analysis of Data

Current Practices and Emerging Trends

Table 1

Frequencies of AI Applications Currently Used in Cataloguing (N=142)

AI Application	n	%
Automated duplicate detection	98	69.0
Auto-generation of MARC records from source data	76	53.5
Machine-learning subject classification	58	40.8
Authority control assistance (name disambiguation)	51	35.9
Linked data entity extraction	37	26.1
Generative AI for summaries/abstracts	24	16.9
Real-time bibliographic enrichment	31	21.8
No AI use in cataloguing	19	13.4

Automated duplicate detection was the most common AI application (69%), followed by auto-generation of MARC records (53.5%). Only 13.4% reported no AI use. Chi-square test revealed no significant difference between North American and European libraries in overall AI adoption ($\chi^2(1) = 2.31$, $p = .128$).

Table 2

Mean Ratings of Emerging Trends in AI for Cataloguing (1=Not at all important, 5=Extremely important)

Emerging Trend	M	SD
AI-assisted authority control	4.21	0.78
Integration with linked data (BIBFRAME)	4.08	0.85
Generative AI for metadata enhancement	3.95	0.92
Real-time enrichment from external sources	3.74	0.98
Automated subject reclassification of legacy data	3.92	0.87

Respondents rated AI-assisted authority control as the most important emerging trend ($M = 4.21$, $SD = 0.78$). Open-ended comments noted interest in using LLMs for "copy cataloguing assistance" and "multi-lingual metadata creation."

Challenges of AI Integration

Table 3

Frequency of Perceived Challenges to AI Integration (N=142, multiple select)

Challenge Category	Specific Challenge	n	%
Technical	Poor legacy data quality	112	78.9
	Integration with ILS/LMS	98	69.0
	Lack of standardized APIs	74	52.1
Organisational	Staff resistance to change	89	62.7
	Insufficient AI training	123	86.6
	Budget for AI tools	101	71.1
Ethical	Algorithmic bias in subject headings	77	54.2
	Privacy of bibliographic data	65	45.8
	Lack of AI governance/audit standards	94	66.2

Insufficient AI training was the most frequently cited challenge (86.6%), followed by poor legacy data quality (78.9%). Ethical concerns about algorithmic bias were noted by 54.2%, but only 45.8% cited privacy—suggesting lower awareness or prioritisation.

Qualitative Thematic Analysis (Interviews, n=12):

Three major themes emerged:

1. **"Garbage in, gospel out"** – Interviewees repeatedly highlighted that AI amplifies existing metadata errors. One head cataloguer noted: "Our 1980s records are inconsistent. AI thinks an edition statement is a publication date." (Participant 7).
2. **"Who teaches the teacher?"** – Lack of in-house expertise was a dominant theme. Several described relying on IT departments or external consultants who lacked domain knowledge of cataloguing rules (RDA, ISBD).
3. **"Trust but verify"** – Participants expressed a need for AI explainability. "If AI suggests a subject heading, I need to know why. Black-box models aren't acceptable for professional accountability" (Participant 3).

Discussion of Findings

The findings reveal that AI integration into cataloguing is progressing but remains uneven, consistent with prior research (Park & Joo, 2022). Regarding current practices emphasise rule-based automation (e.g., duplicate detection) over advanced generative AI, likely due to lower risk and easier validation. The lower adoption of linked data extraction (26.1%) contrasts with the high rated importance of BIBFRAME integration (M = 4.08), suggesting an aspiration–reality gap. This may reflect technical complexity and the nascent state of linked data infrastructure in many libraries. Notably, only 16.9% use generative AI for summaries. Given the rapid evolution of LLMs since 2023, this figure may rise sharply, but interview data revealed caution about hallucination and copyright. Libraries appear to adopt a "wait-and-see" posture, prioritising low-stakes tasks.

The primacy of training deficiencies (86.6%) and legacy data quality (78.9%) indicates that AI readiness is as much about people and data as about technology. This challenges techno-centric narratives that portray AI as a plug-and-play solution. Staff resistance (62.7%) is lower than training deficits, suggesting that opposition is rooted in competency gaps rather than ideological rejection. However, qualitative data revealed subtle anxieties about professional identity, echoing Fitzgerald (2023).

Ethical challenges, particularly algorithmic bias (54.2%), were acknowledged but not universally prioritised. This is concerning because bibliographic metadata shapes access and discoverability; biased subject assignments can perpetuate systemic exclusion (Noble, 2020). The lower concern for privacy (45.8%) may reflect a perception that MARC data is non-sensitive, yet combined with patron usage logs, re-identification risks exist.

Comparison across regions showed no significant difference in adoption rates, contradicting assumptions that south-south concerns would slow AI uptake. However, south-south respondents qualitatively expressed more worry about data sovereignty when using US-based AI cloud services.

Finally, the absence of governance frameworks was identified as a core barrier (66.2%). Without standards for auditing AI accuracy, explainability, and bias, libraries cannot responsibly scale AI beyond experimental phases. This finding aligns with Ross & Taylor (2025) and calls for urgent professional collaboration.

Conclusion and Recommendations

This study provides a systematic, empirical snapshot of AI integration into cataloguing and bibliographic services in academic libraries from 2020 to 2025. Addressing current practices are dominated by routine automation tasks, while emerging trends point toward AI-assisted authority control and linked data enrichment. Regarding, the most formidable challenges are insufficient staff training, poor legacy metadata quality, and absence of ethical governance frameworks. Notably, generative AI remains on the periphery but holds disruptive potential.

Theoretically, the study contributes a typology of AI applications in technical services and a validated challenge taxonomy. Practically, it offers library administrators evidence-based priorities: invest in metadata cleanup before AI adoption, create cross-functional AI literacy programmes, and advocate for community-driven AI audit standards. Limitations include a Western-centric sample, potential self-selection bias (libraries already interested in AI may have over-responded), and the cross-sectional design, which cannot capture longitudinal trends. Future research should examine AI's impact on cataloguing accuracy and user discovery outcomes, and explore non-Western contexts.

Based on the findings, the following recommendations are proposed for academic libraries, professional organisations, and vendors:

1. Develop structured AI training programmes for cataloguing staff, covering both technical skills (using AI tools) and critical evaluation (identifying bias, hallucination).
2. Prioritise legacy data remediation through automated cleanup scripts or crowdsourced correction before implementing AI that depends on those data.
3. Establish local AI governance committees to create policies for testing, deploying, and auditing AI models in bibliographic workflows. Include ethicists and copyright officers.
4. Advocate for interoperable standards (e.g., AI-friendly APIs for ILS, shared model repositories) via organisations like ALA, IFLA, and OCLC.
5. Pilot generative AI in low-risk areas (e.g., drafting abstracts for public domain materials) with mandatory human-in-the-loop review.

6. Participate in collaborative bias auditing by contributing to open datasets of test queries and subject heading assignments.

References

Anderson, R. (2021). Data quality in library legacy systems: Implications for machine learning. *Library Resources & Technical Services, 65*(3), 112–125. https://doi.org/10.5860/lrts.65n3.112

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications.

Breeding, M. (2020). AI and library systems: From hype to reality. *Computers in Libraries, 40*(2), 12–17.

Chen, L. (2024). Large language models in metadata creation: Opportunities and risks. *Journal of Library Metadata, 24*(1), 45–67. https://doi.org/10.1080/19386389.2024.2301298

Creswell, J. W., & Clark, V. L. P. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Fitzgerald, M. (2023). "Will AI replace me?": Cataloguer perceptions of artificial intelligence. *Cataloging & Classification Quarterly, 61*(4), 378–399. https://doi.org/10.1080/01639374.2023.2205955

Gupta, A., & Sharma, R. (2023). Neural networks for automatic subject indexing: A 5-year review. *International Journal on Digital Libraries, 24*(2), 89–104. https://doi.org/10.1007/s00799-023-00345-w

Huang, Y., Zhang, L., & Wang, X. (2024). Knowledge graph-enhanced bibliographic entity linking with BERT. *Journal of Information Science, 50*(1), 212–228. https://doi.org/10.1177/01655515221145678

Kumar, V., & Singh, P. (2022). Interoperability challenges between AI modules and legacy library systems. *Library Hi Tech, 40*(6), 1633–1650. https://doi.org/10.1108/LHT-09-2021-0356

Lee, S., & Kim, J. (2021). AI chatbots in academic libraries: A scoping review. *Reference Services Review, 49*(3/4), 242–260. https://doi.org/10.1108/RSR-03-2021-0012

Martínez-García, A., & López, F. (2023). ORCID and graph neural networks for authority control. *College & Research Libraries, 84*(5), 712–729. https://doi.org/10.5860/crl.84.5.712

Noble, S. U. (2020). *Algorithms of oppression: How search engines reinforce racism* (2nd ed.). New York University Press.

Ochoa, M., & Zhang, Y. (2024). AI continuing education for technical services librarians: A national survey. *Journal of Education for Library and Information Science, 65*(2), 156–174. https://doi.org/10.3138/jelis-2023-0022

Okolo, C. T. (2023). A systematic literature review of AI in library technical services. *Library & Information Science Research, 45*(4), Article 101175. https://doi.org/10.1016/j.lisr.2023.101175

Park, H., & Joo, S. (2022). Automated subject indexing in ARL libraries: Current practices and future directions. *Library Resources & Technical Services, 66*(4), 201–218. https://doi.org/10.5860/lrts.66n4.201

Rodriguez, M. (2025). Generative AI for special

collections metadata: A case study. *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage, 26*(1), 33–51. https://doi.org/10.5860/rbm.26.1.33

Ross, L., & Taylor, N. (2025). Governing algorithmic cataloguing: A proposed framework for academic libraries. *Journal of Academic Librarianship, 51*(1), Article 102433. https://doi.org/10.1016/j.acalib.2024.102433

Smith, J., & Liu, D. (2022). The evolving role of cataloguing in discovery systems. *Library Trends, 70*(3), 421–442. https://doi.org/10.1353/lib.2022.0012

Thompson, C., & White, R. (2024). Real-time bibliographic enrichment using AI web scrapers. *Information Technology and Libraries, 43*(2), 1–19. https://doi.org/10.6017/ital.v43i2.17635