IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double- Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

Review of Artificial Intelligence on Cataloguing, Classification and Metadata Management in Academic Libraries

Aele Navaneetha[1] and Dr. Yogesh Kumar Atri[2]
[1]Research Scholar, [2]Professor
Department of Library And Information Science
Sunrise University, Alwar (Rajasthan)

Abstract: *Artificial Intelligence (AI) is increasingly transforming academic library operations by improving the efficiency, consistency, and scalability of cataloguing, classification, and metadata management. Traditional library processing requires substantial human effort for descriptive cataloguing, subject analysis, classification, authority control, and metadata creation. AI-based technologies, including machine learning, natural language processing, optical character recognition, computer vision, and large language models, provide opportunities to automate or support many of these activities. AI can assist librarians in extracting bibliographic information, generating descriptive metadata, identifying subjects and keywords, recommending classification numbers, detecting duplicate records, linking authority data, and enriching catalogue records.*

In academic libraries, these capabilities can reduce repetitive workloads while improving the discoverability and accessibility of scholarly resources. At the same time, AI adoption introduces challenges related to metadata accuracy, algorithmic bias, interoperability, explainability, data privacy, copyright, standards compliance, and the continuing need for professional judgement.

Keywords: Artificial Intelligence, Academic Libraries, Cataloguing, Classification, Metadata Management, Machine Learning, Natural Language Processing, Library Automation

I. INTRODUCTION

Academic libraries are undergoing significant technological transformation as scholarly communication increasingly incorporates electronic books, digital repositories, research datasets, multimedia resources, institutional publications, open educational resources, and other digital materials. The growth and diversity of these resources have created new demands for efficient organization, description, retrieval, and long-term management of information. Cataloguing, classification, and metadata management remain fundamental library functions because they determine how resources are represented and subsequently discovered by students, researchers, faculty members, and other users. However, conventional approaches often involve detailed intellectual analysis and repetitive data-entry activities that can require considerable time and specialized expertise. The increasing availability of Artificial Intelligence (AI) offers an opportunity to enhance these processes through automated and semi-automated techniques.

AI refers broadly to computational approaches that enable systems to perform tasks associated with human intelligence, including pattern recognition, language processing, prediction, reasoning, and learning from data. In library environments, AI can be applied to bibliographic record creation, subject identification, classification recommendation, entity recognition, authority control, metadata extraction, duplicate detection, and semantic enrichment. Machine learning models can identify patterns in existing catalogue records and use those patterns to recommend metadata for newly acquired resources. Natural language processing can analyze titles, abstracts, tables of contents, and full-text documents to identify subjects, concepts, keywords, and entities. Optical character recognition can convert scanned materials into machine-readable text, while computer vision can assist in interpreting images and other visual materials.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double- Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

The relevance of AI is particularly significant for academic libraries because their collections are extensive, multidisciplinary, and continuously expanding. Research materials frequently contain specialized terminology, multiple languages, complex authorship structures, institutional affiliations, and relationships between publications. AI can potentially support the management of these complexities by connecting information across different metadata elements and resources. It can also contribute to improved discovery by generating richer metadata and semantic relationships than may be possible through minimal descriptive records.

Nevertheless, AI-based cataloguing should not be viewed simply as an automation exercise. Bibliographic description and classification involve professional standards, contextual interpretation, and decisions about how resources should be represented to particular communities. Machine-generated metadata may contain factual errors, inappropriate subject terms, inconsistent classifications, or hidden biases present in training data. Consequently, academic libraries need appropriate validation mechanisms, metadata policies, ethical guidelines, and professional oversight. The future role of AI therefore appears to be complementary rather than purely substitutive, combining computational efficiency with librarians' domain knowledge and intellectual judgement.

II. ARTIFICIAL INTELLIGENCE IN CATALOGUING AND BIBLIOGRAPHIC RECORD CREATION

Cataloguing is one of the library activities in which AI can provide substantial assistance. Conventional cataloguing requires librarians to examine resources, identify bibliographic elements, construct descriptions, determine access points, establish authority relationships, and create or edit machine-readable records. For large academic collections, these activities can become time-consuming, particularly when libraries acquire substantial numbers of electronic resources and digitized materials. AI can automate portions of this workflow by extracting information from digital objects and converting it into structured metadata.

Natural language processing is particularly useful for analyzing textual information. An AI system can process a title, abstract, table of contents, introduction, or complete document and identify potentially important metadata elements. Named-entity recognition can identify personal names, organizations, geographical locations, dates, and other entities. These entities can subsequently be matched with authority files or knowledge bases. AI can also assist in identifying missing metadata and recommending values based on patterns observed in comparable records.

Optical character recognition further expands the possibilities for cataloguing historical and digitized collections. Scanned books, manuscripts, theses, newspapers, and archival materials can be converted into machine-readable text, after which AI techniques can extract names, subjects, dates, and other descriptive information. This can substantially reduce the amount of manual transcription required for retrospective conversion projects.

AI can also support copy cataloguing by identifying similar or duplicate bibliographic records. A system may compare ISBNs, titles, authors, publication details, abstracts, and other attributes to locate existing records. Instead of creating a new record from scratch, librarians can review the recommended record and modify it where necessary. Such systems can improve workflow efficiency while preserving professional control over the final record.

Large language models have created additional possibilities for metadata generation because they can summarize texts, identify concepts, generate keywords, and transform unstructured information into structured descriptions. However, their outputs require careful verification because generative AI may produce inaccurate or unsupported information. In academic library environments, therefore, AI-generated cataloguing should preferably operate within controlled workflows based on authoritative data sources and clearly defined metadata standards.

III. ARTIFICIAL INTELLIGENCE IN CLASSIFICATION AND SUBJECT ANALYSIS

Classification is another area where AI can assist academic librarians. Library classification involves analyzing the intellectual content of a resource and assigning an appropriate class number or subject category. Established systems such as the Dewey Decimal Classification and Library of Congress Classification provide structured frameworks, but assigning precise classifications may require considerable subject expertise. AI can learn from existing classified collections and generate recommendations for new resources.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double- Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

Machine learning algorithms can be trained using historical catalogue records containing titles, abstracts, subject headings, and classification numbers. Once trained, a model can identify patterns between textual characteristics and classification decisions. When a new document is introduced, the system analyzes its content and proposes one or more likely classification categories. Librarians can then accept, modify, or reject the recommendation.

Natural language processing can also support subject analysis. Instead of relying exclusively on the title of a resource, AI can analyze abstracts and full text to identify the major concepts discussed in a document. This is particularly useful for scholarly literature where titles may be ambiguous or highly specialized. Semantic analysis can identify relationships between concepts and suggest subject headings or descriptors.

AI-assisted classification can be especially valuable for institutional repositories and digital collections. Academic institutions generate large quantities of theses, dissertations, research papers, conference materials, datasets, and technical reports. Automatically recommending subject categories can make these resources more searchable and consistently organized.

AI Application	Function in Classification	Potential Benefit	Human Role
Machine learning	Recommends classification numbers	Faster processing	Validate recommendations
NLP	Identifies concepts and subjects	Improved subject analysis	Review conceptual relevance
Semantic analysis	Detects relationships between concepts	Better discovery	Check contextual accuracy
Text classification	Assigns documents to categories	Scalable processing	Correct ambiguous cases
Similarity analysis	Finds comparable classified resources	Supports consistent decisions	Select appropriate precedent
Generative AI	Suggests subject descriptions	Rapid metadata enrichment	Verify generated content

Despite these advantages, AI classification systems can reproduce biases contained in historical catalogue data. A model trained on existing records may learn older terminology or classification practices that no longer reflect contemporary scholarly perspectives. It may also struggle with interdisciplinary works that legitimately belong to multiple subject areas. Therefore, AI recommendations should be considered decision-support outputs rather than unquestionable classifications.

IV. AI-DRIVEN METADATA MANAGEMENT AND ENRICHMENT

Metadata management involves the creation, organization, maintenance, exchange, and preservation of structured information describing library resources. In modern academic libraries, metadata exists across integrated library systems, discovery services, institutional repositories, digital libraries, research information systems, and specialized databases. Maintaining consistency across these environments is increasingly challenging. AI can support metadata management by identifying inconsistencies, enriching records, detecting relationships, and recommending corrections.

One important application is automatic metadata extraction. AI systems can analyze digital documents and identify authors, titles, affiliations, abstracts, keywords, dates, funding information, references, and other relevant elements. These data can be mapped to appropriate metadata fields. Such automation can accelerate the processing of institutional repository materials and digitized collections.

AI can also contribute to metadata normalization. Records obtained from different sources may contain variations in names, publication information, dates, subject terminology, or institutional names. Machine learning and entity-resolution techniques can identify records that refer to the same entity even when their textual representations differ. This improves consistency and supports more effective information retrieval.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double- Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

Another important application is metadata enrichment. AI can add keywords, summaries, subject concepts, related resources, citation relationships, geographic information, and semantic links to existing records. Richer metadata can improve discovery by allowing users to search resources through multiple conceptual pathways.

Metadata Management Activity	AI Technique	Expected Contribution
Metadata extraction	NLP and machine learning	Automatic identification of descriptive elements
Entity recognition	Named-entity recognition	Identification of persons, organizations and places
Metadata normalization	Entity matching	Greater consistency across records
Duplicate detection	Similarity algorithms	Reduction of redundant records
Keyword generation	NLP and language models	Improved resource discoverability
Metadata enrichment	Semantic analysis	Addition of contextual relationships
Quality control	Anomaly detection	Identification of incomplete or inconsistent records
Authority control	Entity resolution	Better consistency of names and entities

The effectiveness of AI-driven metadata management depends strongly on interoperability with established library standards and systems. AI-generated information must be mapped appropriately to formats and standards used by libraries. It should also preserve provenance so that librarians can determine how a metadata element was created, modified, or recommended. Transparent provenance is particularly important when AI is used to modify authoritative bibliographic information.

V. BENEFITS, CHALLENGES AND ETHICAL CONSIDERATIONS

The adoption of AI in academic library technical services can generate several important benefits. The most immediate advantage is increased efficiency. AI can process large volumes of records faster than manual workflows, allowing librarians to devote more time to complex cataloguing, collection development, research support, and user services. Automation can also reduce repetitive data-entry tasks and potentially improve consistency across large collections.

AI can support scalability as academic libraries increasingly manage digital resources. A human-only workflow may become difficult when thousands of documents, datasets, and institutional publications require metadata. AI-assisted processing can help libraries manage these expanding collections without proportionally increasing manual workload.

Improved discovery is another major benefit. Richer and more consistent metadata can enable users to locate resources through keywords, concepts, entities, subjects, and semantic relationships. AI can also identify relationships between resources that might not be explicitly represented in conventional cataloguing records.

However, several challenges must be addressed. Accuracy is a central concern because AI-generated metadata may contain errors. Generative systems may create plausible but incorrect information, while classification algorithms can make inappropriate recommendations for interdisciplinary or unusual materials. Algorithmic bias is another concern because models learn from existing data. If historical catalogue records contain biases or outdated terminology, automated systems may reproduce them.

Privacy and data governance are also important. Academic library materials may include unpublished research, student theses, sensitive datasets, or restricted institutional information. Libraries must carefully determine what information can be processed by external AI services and how that information is stored and used.

Copyright and intellectual-property considerations require attention when AI systems analyze full-text resources or generate derivative metadata. Academic libraries should establish clear policies governing the use of AI systems and ensure compliance with applicable legal and institutional requirements.

Professional ethics are equally important. Librarians have traditionally been responsible for ensuring that information is organized fairly, accurately, and accessibly. Excessive dependence on automated systems could weaken professional oversight. Consequently, a human-in-the-loop model is preferable. AI should generate recommendations and perform repetitive operations, while librarians retain responsibility for validation, correction, policy interpretation, and final decision-making.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double- Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

VI. CONCLUSION

The future of AI in academic library technical services is likely to involve increasingly integrated systems that combine machine learning, natural language processing, knowledge graphs, computer vision, and generative AI. Instead of functioning as isolated automation tools, these technologies may become integrated into library management platforms and discovery environments. AI systems may continuously analyze catalogue records, identify inconsistencies, recommend metadata improvements, and connect resources through semantic relationships.

Future developments are also likely to strengthen multilingual cataloguing and classification. Academic libraries serve increasingly international research communities, and AI-based language technologies can assist in translating metadata, identifying equivalent concepts across languages, and improving access to multilingual resources. Knowledge graphs may further enhance discovery by connecting authors, institutions, research topics, publications, datasets, and related entities.

Another important direction is explainable AI. Librarians need to understand why a system recommends a particular classification number, subject heading, or metadata value. Explainable systems can provide evidence or reasoning that allows professionals to evaluate recommendations efficiently. This can increase trust and make AI-supported workflows more accountable.

Academic libraries should therefore approach AI adoption strategically rather than simply pursuing maximum automation. Before implementation, libraries should evaluate data quality, metadata standards, interoperability, privacy, security, bias, cost, and staff training requirements. Continuous evaluation should be incorporated into AI-supported workflows to measure accuracy and identify systematic errors.

Overall, Artificial Intelligence has considerable potential to transform cataloguing, classification, and metadata management in academic libraries. AI can accelerate bibliographic processing, support classification decisions, automate metadata extraction, identify duplicate and inconsistent records, and enrich discovery systems. Nevertheless, these advantages are accompanied by concerns about accuracy, bias, privacy, transparency, professional responsibility, and standards compliance. The most sustainable model is therefore not the replacement of librarians but collaboration between human expertise and intelligent computational systems. Librarians will increasingly function as metadata strategists, quality controllers, ethical overseers, and evaluators of AI-generated information. When implemented responsibly, AI can strengthen rather than diminish the fundamental mission of academic libraries: providing reliable, organized, discoverable, and equitable access to scholarly knowledge.

REFERENCES

1. Asemi, A., Ko, A., & Nowkarizi, M. (2021). Intelligent libraries: A review on expert systems, artificial intelligence, and robot. *Library Hi Tech, 39*(2), 412–434.
2. Cox, A. M. (2021). Exploring the impact of artificial intelligence, machine learning and automation on information professionals. *Journal of the Association for Information Science and Technology, 72*(2), 189–202.
3. Cox, A. M., Pinfield, S., & Rutter, S. (2019). The intelligent library: Thought leaders' views on the likely impact of AI on academic libraries. *Library Hi Tech, 37*(3), 418–435.
4. Fourie, I. (2018). The role of artificial intelligence in libraries: Current trends and future possibilities. *Library Hi Tech, 36*(4), 680–693.
5. Lo, L. S. (2023). The rise of generative artificial intelligence and its impact on academic libraries. *Library Hi Tech, 41*(6), 1901–1917.
6. Momeni, R., & Taghizadeh, M. (2021). Artificial intelligence and its applications in libraries: A systematic review. *Library Philosophy and Practice, 2021*, 1–18.
7. Ocholla, D. N., & Ocholla, L. (2020). Readiness of academic libraries for artificial intelligence adoption: A review of emerging issues. *South African Journal of Libraries and Information Science, 86*(2), 1–12.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double- Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, July 2026
Impact Factor: 8.2

8. Smith, M. S. (2020). Artificial intelligence in libraries: Applications, challenges, and opportunities. *Library Management, 41*(6/7), 435–447.
9. Wang, Y., & Alexander, C. (2020). Big data and artificial intelligence in academic libraries: Emerging applications and challenges. *The Electronic Library, 38*(5), 1–15.
10. Zhang, Y., & Ma, X. (2022). Artificial intelligence applications in academic libraries: A systematic review. *Journal of Academic Librarianship, 48*(5), 102566.
11. Breeding, M. (2019). The future of library resource discovery: Artificial intelligence and machine learning. *Computers in Libraries, 39*(5), 16–19.
12. Golub, K., & Liu, Y.-H. (2017). Automated subject indexing of text. *Journal of Documentation, 73*(4), 769–790.
13. Golub, K., Lykke Nielsen, M., & Liu, Y.-H. (2016). Automated subject indexing of text. *Journal of Documentation, 72*(4), 634–657.
14. Hider, P., & Liu, Y.-H. (2013). Automated classification in digital libraries: A review of methods and applications. *Library Hi Tech, 31*(4), 643–657.
15. Mai, J.-E. (2016). Metadata. *Knowledge Organization, 43*(5), 354–361.
16. Park, J.-R. (2009). Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging & Classification Quarterly, 47*(3/4), 213–228.
17. Zeng, M. L., & Qin, J. (2016). *Metadata*. Facet Publishing.
18. Svenonius, E. (2000). *The intellectual foundation of information organization*. MIT Press.
19. Taylor, A. G., & Joudrey, D. N. (2018). *The organization of information* (4th ed.). Libraries Unlimited.
20. Bawden, D., & Robinson, L. (2022). The dehumanization of information: AI, automation, and the future of information work. *Journal of the Association for Information Science and Technology, 73*(10), 1427–1439.