

Library and Information Science: Theory and Digital Trends

ISBN: 978-93-49938-87-8 | Year: 2026 | pp: 01 - 17 |

Transforming Knowledge Organization, Information Retrieval and Library Services in the Digital Age

¹Divyansh Mishra

²Rajesh Kumar Mishra

³Rekha Agarwal

¹Department of Artificial Intelligence and Data Science, Jabalpur Engineering College, Jabalpur (MP)

²ICFRE-Tropical Forest Research Institute (Ministry of Environment, Forests & Climate Change, Govt. of India), P.O. RFRC, Mandla Road, Jabalpur, MP-482021, India

³Government Science College, Jabalpur, MP, India- 482 001

Email: rajeshkmishra20@gmail.com

Article DOI Link: <https://zenodo.org/uploads/20833466>

DOI: 10.5281/zenodo.20833466

Abstract

This chapter provides a comprehensive examination of Artificial Intelligence (AI) and Machine Learning (ML) as transformative forces within Library and Information Science (LIS). Beginning with a historical overview of AI adoption in library contexts, the chapter progresses through core ML methodologies—including supervised and unsupervised learning, natural language processing, and deep learning architectures—before analysing their applied impact on cataloguing, classification, reference services, collection development, and digital preservation. The chapter addresses the critical ethical dimensions of algorithmic bias, data privacy, and the evolving professional identity of library and information professionals. Particular attention is given to large language models and generative AI as emergent technologies reshaping scholarly communication and information literacy. The chapter concludes with a forward-looking synthesis of strategic imperatives for LIS institutions navigating an AI-mediated information environment.

Keywords: Artificial Intelligence, Machine Learning, Natural Language Processing, Knowledge Organization, Information Retrieval, Algorithmic Bias, Digital Libraries, Generative AI, LIS Profession

Introduction

The convergence of Artificial Intelligence (AI) and Machine Learning (ML) with Library and Information Science (LIS) represents one of the most consequential paradigmatic shifts in the history of organised knowledge. Libraries—as institutions fundamentally concerned with the acquisition, organisation, preservation, and dissemination of recorded knowledge—find themselves at an extraordinary inflection point, where centuries-old intellectual traditions collide with computational methodologies of unprecedented power and scope [1]. Historically, libraries have never been passive repositories; they have always been active participants in the social construction of knowledge. From the cataloguing reforms introduced by Charles Cutter in the late nineteenth century to the machine-readable cataloguing (MARC) standards developed at the Library of Congress in the 1960s, the profession has continuously adapted its intellectual frameworks to accommodate new technological paradigms [2]. What distinguish the current transformation, however, are both its velocity and its epistemic depth: AI systems do not merely automate pre-existing human procedures but increasingly perform cognitive functions—semantic inference, contextual disambiguation, predictive reasoning—that were previously understood to be the exclusive province of human intellect [3]. This chapter addresses the multifaceted relationship between AI/ML and LIS at the postgraduate and research level, presupposing familiarity with foundational information science concepts. Section 2 traces the intellectual genealogy of AI in library contexts. Section 3 provides a rigorous exposition of core ML methodologies relevant to LIS practice and research. Sections 4 through 7 examine applied domains in depth. Section 8 engages with the ethical and professional implications of AI adoption. Section 9 addresses the emergent phenomenon of large language models (LLMs) and generative AI. The chapter concludes in Section 10 with a synthesis of strategic imperatives for the profession.

Historical Context: AI in Library and Information Science

• Early Automation and Expert Systems (1960s–1990s)

The intellectual antecedents of AI in LIS can be traced to the emergence of information retrieval (IR) as a formal discipline in the 1950s and 1960s. Gerard Salton's development of the SMART system at Cornell University introduced the vector space model, establishing a mathematical framework for document-query similarity that remains foundational to contemporary search algorithms [4]. Concurrently, the MEDLINE system at the National Library of Medicine pioneered large-scale automated indexing, demonstrating the feasibility of machine-assisted

bibliographic control at institutional scale [5]. The 1980s witnessed the proliferation of expert systems in library applications. Systems such as CANSEARCH and the Reference Expert aimed to encode the knowledge of reference librarians in rule-based inference engines, enabling automated mediation between user queries and bibliographic databases [6]. These early systems, while intellectually significant, were ultimately constrained by the brittleness of knowledge representation in purely symbolic AI: they could not generalise beyond their explicitly programmed domains and were incapable of learning from user interactions.

- **The Statistical Turn and the World Wide Web (1990s–2000s)**

The statistical revolution in natural language processing, catalyzed by the availability of large corpora and increasing computational power, fundamentally reoriented AI research away from hand-crafted rules toward data-driven models. The seminal work of Church and Hanks on word association norms, and subsequent developments in probabilistic language modelling, provided LIS researchers with new tools for automated indexing, query expansion, and cross-language information retrieval [7]. The emergence of the World Wide Web in the early 1990s simultaneously created an unprecedented information management challenge and an extraordinary data resource for training statistical models. Libraries were compelled to extend their organizational frameworks to encompass digital objects, leading to the development of metadata standards such as Dublin Core (1995) and the Encoded Archival Description (EAD) schema. Search engines such as AltaVista and, subsequently, Google demonstrated the commercial viability of large-scale automated information organization, fundamentally reshaping public expectations of search and retrieval [8].

- **The Deep Learning Renaissance (2010s–Present)**

The publication of Hinton et al.'s landmark work on deep belief networks in 2006, followed by the breakthrough performance of deep convolutional neural networks in the ImageNet Large Scale Visual Recognition Challenge in 2012, inaugurated a new era in AI capability. [9] The subsequent development of attention mechanisms and transformer architectures—culminating in Google's BERT (2018) and OpenAI's GPT series—produced language models capable of nuanced semantic understanding that approached, and in many benchmarks exceeded, human-level performance on specific tasks [10]. For LIS, this deep learning renaissance has translated into qualitatively new possibilities: neural machine translation enabling multilingual access to collections; image recognition facilitating visual metadata generation for photograph archives; and conversational AI enabling more sophisticated virtual reference services. The challenge for the profession is no longer primarily technical feasibility but rather the critical evaluation, ethical deployment, and intellectual governance of these powerful systems [11].

Core Machine Learning Methodologies Relevant to LIS

- **Supervised Learning**

Supervised learning encompasses algorithms that learn a mapping function from labelled input-output pairs, enabling the classification or regression of unseen instances. In LIS contexts, supervised learning is the dominant paradigm for automated cataloguing and subject classification, where training data typically consists of bibliographic records paired with human-assigned subject headings or Dewey Decimal classifications [12]. Support vector machines (SVMs), random forests, and gradient boosting algorithms have been extensively applied to document classification tasks. More recently, fine-tuned transformer models—particularly domain-adapted variants of BERT trained on scholarly text—have achieved classification accuracies that substantially reduce, though do not eliminate, the need for human cataloguing intervention [13]. A critical methodological consideration is the quality and representativeness of training data: classifiers trained on historical cataloguing data will inherit and potentially amplify the subject biases embedded in those records, a problem examined in detail in Section 8.

- **Unsupervised Learning and Topic Modelling**

Unsupervised learning algorithms identify latent structures in data without recourse to pre-labelled examples. Topic modelling, particularly Latent Dirichlet Allocation (LDA) introduced by Blei, Ng, and Jordan in 2003, has been widely adopted in LIS for the exploration of large document collections [14]. LDA models a corpus as a mixture of latent topics, each characterised by a probability distribution over vocabulary terms, enabling the discovery of thematic structures that may not align with existing controlled vocabulary schemas. Applications in LIS include the analysis of historical newspaper archives, the identification of intellectual themes in institutional repositories, and the exploration of interdisciplinary connections within research literature. Neural topic models, which leverage word embedding representations, offer improved coherence over classical LDA and are increasingly favoured in contemporary research [15].

- **Natural Language Processing (NLP)**

Natural language processing constitutes perhaps the most consequential sub domain of AI for LIS, given that the objects of library organization are predominantly textual. The NLP pipeline relevant to LIS encompasses tokenization, part-of-speech tagging, named entity recognition (NER), dependency parsing, semantic role labelling, co reference resolution, and, at the discourse level, summarization and machine translation [16]. Named entity recognition is particularly significant for archival and bibliographic applications, enabling the automatic identification and linking of persons, organizations, places, and dates within textual records. When combined with authority files such as the Library of Congress Name Authority File

(LCNAF) or the Virtual International Authority File (VIAF), NER enables the semi-automated construction of linked data representations that support faceted search and knowledge graph applications [17].

- **Neural Networks, Embeddings, and Transformers**

Word embeddings, introduced in their modern form by Mikolov et al.'s Word2Vec (2013) and subsequently refined by GloVe and fastText, represent words as dense vectors in a continuous semantic space, capturing distributional similarity relationships [18]. For LIS, embeddings have transformed query expansion, synonym generation, and cross-language retrieval. Contextualized embeddings, produced by transformer architectures such as BERT and RoBERTa, further refine this representation by conditioning word meaning on surrounding context, resolving polysemy that was intractable for static embedding approaches [19]. The transformer's self-attention mechanism, which enables the model to dynamically weight the relevance of all tokens in a sequence relative to one another, constitutes a fundamental architectural advance over recurrent neural networks for long-range dependency modelling. For LIS researchers, understanding the transformer architecture is essential for evaluating claims about model capability, diagnosing failure modes, and critically assessing the interpretability limitations of black-box neural systems [20].

Automated Cataloguing and Subject Classification

- **The Bibliographic Control Challenge**

Bibliographic control—the systematic description, identification, and organisation of information resources—has constituted the intellectual core of professional librarianship since the nineteenth century. The transition to digital environments has dramatically amplified the volume of materials requiring cataloguing: institutional repositories, open access journals, digitised heritage collections, and grey literature repositories collectively represent cataloguing backlogs of a scale that renders manual treatment economically untenable [21]. The Library of Congress Cooperative Cataloging Program and OCLC's WorldCat represent significant institutional responses to this challenge, distributing cataloguing labour across member institutions and enabling record sharing. AI-assisted cataloguing represents the next evolutionary step, promising to reduce the per-record cost of cataloguing while maintaining or enhancing intellectual quality [22].

- **Machine Learning Approaches to Subject Heading Assignment**

Automated subject heading assignment is among the most extensively researched applications of ML in LIS. The Annif toolkit, developed at the National Library of Finland, provides an open-source platform for automated subject indexing using an ensemble of ML models, including TF-IDF baselines, neural network classifiers,

and transformer-based models [23]. Evaluations conducted on the National Bibliography demonstrate that transformer-based Annif models achieve F1 scores competitive with, though not consistently superior to, professional human indexers for high-frequency subject headings, with performance degrading significantly for rare and highly specialised topics. The Dewey Decimal Classification (DDC) assignment problem presents particular challenges for automated systems, given the hierarchical, polyhierarchical, and schedule-notation complexity of the DDC scheme. Research by Mäkinen et al. (2022) using fine-tuned SciBERT models achieved 72.3% accuracy at the DDC class level for scientific monographs, with accuracy declining to 48.7% at the division level—figures that highlight both the promise and the current limitations of automated classification [24].

- **Linked Data, BIBFRAME, and Knowledge Graphs**

The transition from MARC to BIBFRAME (Bibliographic Framework Initiative) as the foundational metadata standard for North American libraries represents an alignment of bibliographic practice with semantic web and linked data principles. Within the BIBFRAME model, bibliographic entities—Works, Instances, and Items—are represented as nodes in a knowledge graph, with properties and relationships encoded as RDF triples [25]. ML contributes to linked data workflows through entity linking, the task of aligning named entities in metadata records with canonical identifiers in authority files and knowledge bases such as Wikidata and DBpedia. Neural entity linking systems leverage contextualized embeddings to disambiguate name variants, detect co referents, and identify novel entities requiring authority record creation. The potential for AI to accelerate the construction of richly interlinked bibliographic knowledge graphs represents a significant opportunity for enhanced discoverability and scholarly navigation [26-27].

AI-Enhanced Information Retrieval Systems

- **From Boolean to Neural Retrieval**

Classical information retrieval systems, grounded in the Boolean and vector space models, rely on lexical matching between query terms and document representations. While robust and transparent, these approaches are fundamentally limited by the vocabulary mismatch problem: users frequently employ query terms that differ from those used by document authors, resulting in relevant documents being missed [28]. Neural information retrieval addresses vocabulary mismatch through dense vector representations of queries and documents in a shared semantic space, enabling similarity computation that transcends lexical identity. Bi-encoder architectures, such as Dense Passage Retrieval (DPR), encode queries and documents independently into dense vectors, enabling efficient approximate nearest-neighbour search over large corpora. Cross-encoder architectures, which

concatenate query-document pairs for joint encoding, achieve superior relevance estimation but at substantially higher computational cost, making them more appropriate for re-ranking than initial retrieval [29].

- **Retrieval-Augmented Generation in Library Discovery**

Retrieval-Augmented Generation (RAG), introduced by Lewis et al. in 2020, combines dense retrieval with generative language modelling to produce responses grounded in retrieved evidence [30]. In library discovery contexts, RAG systems can retrieve relevant passages from institutional repositories or licensed databases and generate synthesised responses with explicit provenance citations. Early implementations in academic library reference services have demonstrated promising capacity for handling complex research queries, though critical limitations remain in the handling of temporal currency, disciplinary depth, and citation accuracy [31].

- **Personalisation and Contextual Recommendation**

Collaborative filtering and content-based recommendation algorithms have been applied to library discovery for over two decades, with systems such as bX Recommender (now part of Ex Libris Primo) leveraging aggregated usage data to surface contextually relevant resources. Contemporary neural recommendation systems, including graph neural network-based approaches that model complex user-item interaction patterns, represent a significant advance in recommendation accuracy [32]. A substantive tension exists between the potential for personalized recommendation to enhance resource discovery and the risk of creating filter bubbles that constrain scholarly exposure to conceptually adjacent or epistemically challenging material. Library and information scientists are uniquely positioned to interrogate this tension, given the profession's foundational commitment to intellectual freedom and breadth of access [33].

Intelligent Reference Services and Virtual Assistants

- **The Evolution of Digital Reference**

Digital reference—the provision of reference services through electronic channels—emerged in the 1990s and evolved through email-based, chat-based, and video-based modalities. The integration of AI into this service domain introduces new modalities of interaction and fundamentally reconfigures the human-mediated reference transaction [34]. Chatbot-based virtual reference systems, ranging from rule-based FAQ bots to sophisticated LLM-powered conversational agents, have been widely deployed in academic and public libraries. Evaluations of early chatbot deployments revealed significant limitations in the handling of complex, multi-turn queries, sensitivity to query phrasing, and inability to express calibrated

uncertainty—limitations that contributed to user frustration and, in some cases, the provision of substantively incorrect information [35].

- **Large Language Models in Reference Interactions**

The integration of large language models into reference services represents both a significant capability advance and a novel risk profile. LLM-powered reference agents can engage in extended multi-turn dialogue, decompose complex research queries, explain database search strategies, interpret citation formats, and synthesize information from multiple sources—tasks that approximate aspects of professional reference work [36]. However, the phenomenon of LLM hallucination—the generation of plausible but factually incorrect information, including fabricated citations—constitutes a fundamental challenge for reference applications where information accuracy is paramount. Mitigation strategies include RAG architectures that ground model outputs in verified sources, explicit uncertainty quantification mechanisms, and hybrid human-AI workflows that route complex queries to human librarians [37].

AI in Digital Preservation and Collection Development

- **Automated Preservation Risk Assessment**

Digital preservation—the active management of digital objects to ensure their long-term accessibility and authenticity—has become a central operational challenge for libraries, archives, and digital repositories. The complexity of digital object formats, the pace of technological obsolescence, and the scale of contemporary digital collections necessitate AI-assisted approaches to preservation risk assessment and format migration planning [38]. ML models trained on the PRONOM technical registry and institutional preservation records have been developed to predict format obsolescence risk, identify objects requiring immediate format migration, and recommend target formats that balance authenticity, accessibility, and long-term stability. The JHOVE (JSTOR/Harvard Object Validation Environment) and DROID (Digital Record Object Identification) tools provide foundational technical metadata that can serve as features for such predictive models [39].

- **Computer Vision for Digitised Heritage Collections**

Computer vision applications have substantially advanced the accessibility and discoverability of digitised heritage collections. Optical Character Recognition (OCR) has been a library application since the 1970s, but the transition to neural network-based OCR engines—particularly Tesseract 4.0's LSTM engine and the open-source Kraken system optimised for historical documents—has dramatically improved accuracy for degraded, non-standard, and non-Latin scripts [40]. Beyond OCR, deep learning models trained on annotated historical photograph collections enable automated generation of descriptive metadata, identification of depicted

persons through facial recognition (subject to significant ethical constraints explored in Section 8), detection of depicted locations through scene understanding, and classification of image content according to iconographic schemas such as Iconclass. The Bibliothèque nationale de France's Gallica platform and the Europeana initiative represent leading examples of large-scale computer vision deployment in digitised heritage contexts [41-42].

• **Intelligent Collection Development**

Evidence-Based Acquisitions (EBA) and Demand-Driven Acquisitions (DDA) have transformed collection development practice in academic libraries, shifting selection decisions from librarian expertise to demonstrated user demand. ML models that integrate circulation data, interlibrary loan patterns, citation analysis, and open access availability signals enable more sophisticated collection gap analysis and prospective acquisition prioritization [43]. Text mining of research output from affiliated institutions—leveraging NLP pipelines applied to preprints, theses, and published articles—enables libraries to anticipate emerging research needs and identify collection gaps before they become service failures. Such approaches require careful governance frameworks to prevent the replication of existing collection biases and to ensure equitable coverage of interdisciplinary and emergent research areas [44].

Ethical Dimensions and the Professional Identity of the Information Scientist

• **Algorithmic Bias in Bibliographic Systems**

The deployment of ML systems in LIS is not ethically neutral. Algorithms trained on historical bibliographic data inherit and can amplify the structural biases embedded in those records—biases that reflect historical patterns of exclusion, marginalisation, and epistemic violence. The critical analysis of Library of Congress Subject Headings (LCSH) by Sanford Berman and, more recently, the work of Hope Olson on the bias of the universal, have established a rich theoretical framework for understanding how knowledge organisation systems encode particular worldviews [45-46]. When ML classifiers are trained on these historically biased records, they risk reifying and perpetuating problematic representational patterns at machine scale and speed. Research by Noble (2018) on algorithmic oppression in search systems, and by Bates (2017) on the epistemological assumptions embedded in information architecture, provides essential critical context for LIS professionals evaluating AI system deployments [47-48].

Table 7.1: Ethical Risk Dimensions in AI/ML Library Applications

Ethical Risk Dimension	Relevant LIS Application Domain
Training data bias replication	Automated subject classification, cataloguing

Facial recognition misidentification	Heritage photograph description
Filter bubble creation	Personalised recommendation systems
LLM hallucination / citation fabrication	AI-assisted reference services
User privacy erosion through behavioural profiling	Discovery system personalisation
Opacity of algorithmic decision-making	Acquisition prioritisation, deselection
Digital divide amplification	AI-mediated access and search
Over-reliance reducing professional critical judgment	All AI-assisted workflows

Privacy, Surveillance, and Reader Confidentiality

Libraries have historically maintained strong commitments to reader privacy, grounded in the recognition that intellectual freedom requires the freedom to explore ideas without surveillance. The data infrastructures required to train and operate personalised AI systems—detailed logs of search queries, resource access patterns, and interaction histories—represent a fundamental tension with these commitments [49]. The General Data Protection Regulation (GDPR) in the European context and various national data protection frameworks provide a minimum legal floor for user data handling, but LIS professionals and researchers are called to advocate for ethical standards that exceed legal requirements. The American Library Association's Privacy Policy for Library Websites and the IFLA Statement on Privacy in the Library Environment offer normative frameworks, though their application to AI-mediated services requires substantial elaboration [50].

Interpretability, Accountability, and Algorithmic Transparency

The opacity of deep learning models—their status as black boxes resistant to straightforward human interpretation—raises fundamental accountability questions when AI systems are deployed in consequential library decisions such as collection deselection, access restriction, or reference query routing. Explainable AI (XAI) techniques, including LIME (Locally Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations), offer post-hoc interpretability mechanisms, but these are approximations that do not provide genuine mechanistic insight into model reasoning [51].

The Evolving Professional Identity of the Information Scientist

The automation of cataloguing, indexing, and reference tasks raises legitimate questions about the future professional identity of library and information scientists. A technocentric reading might anticipate a straightforward displacement of human expertise; a more nuanced analysis reveals a transformation and repositioning of

professional labour rather than its elimination [52]. The critical evaluation of AI system outputs, the governance of algorithmic bias, the design of ethical AI deployment frameworks, the mediation between technological capability and community need, and the exercise of intellectual judgment in edge cases that confound automated systems—these functions require the deep domain expertise, ethical sensibility, and contextual understanding that remain distinctively human professional capacities. The emerging role of the 'AI-informed information scientist' requires competencies in both computational literacy and the humanistic, social-scientific, and ethical traditions of LIS scholarship [53].

Large Language Models and Generative AI: Emergent Challenges and Opportunities

Architectures and Capabilities of LLMs

Large language models, pre-trained on corpora of unprecedented scale using self-supervised objectives such as masked language modelling and next-token prediction, have achieved qualitative capability advances that have substantially altered the landscape of AI applicability in knowledge work. Models such as GPT-4, Claude, Gemini, and their open-source counterparts—LLaMA, Mistral, and Falcon—demonstrate capacities for multi-step reasoning, long-context comprehension, code generation, and instruction following that were not anticipated by pre-2020 scaling projections [54]. For LIS researchers, the critical assessment of LLM capability claims requires familiarity with evaluation methodologies and their limitations. Benchmark performance on curated task datasets frequently overstates real-world capability due to test set contamination, distributional shift, and the tendency of benchmarks to measure narrowly defined skills rather than robust general competence [55].

Implications for Scholarly Communication and Information Literacy

The integration of LLMs into scholarly communication workflows—manuscript drafting, literature review synthesis, abstract generation, peer review assistance—raises fundamental questions for libraries as institutional stewards of the scholarly record. The provenance, authenticity, and intellectual responsibility of AI-assisted scholarly outputs are contested across disciplines, with emerging norms varying substantially between fields [56]. For information literacy practice and instruction, LLMs present both pedagogical opportunities and challenges. The capacity of LLMs to generate plausible but unverified information creates new cognitive traps for users who lack the critical evaluation skills to distinguish AI-generated confabulation from grounded knowledge claims. Frameworks such as the ACRL Framework for Information Literacy for Higher Education require substantive revision to address the epistemological challenges posed by LLM-mediated information access [57].

Generative AI and Archival Description

Archival institutions face particular challenges in leveraging generative AI given the primacy of archival principles—provenance, original order, and the evidential integrity of the archival bond—that resist the fluid, probabilistic nature of LLM outputs. Nevertheless, experimental applications of LLMs to the generation of scope and content notes, biographical notes, and finding aid narrative from unstructured archival metadata demonstrate meaningful productivity gains with acceptable quality for draft generation subject to professional review [58].

Strategic Imperatives for LIS Institutions and the Research Agenda

The preceding analysis suggests several strategic imperatives for LIS institutions navigating the AI transformation, and a corresponding research agenda for the scholarly community.

Strategic Imperatives for LIS Institutions

- **Invest in AI Literacy Infrastructure:** Develop institutional competency frameworks for AI literacy at all levels of the profession, encompassing both technical understanding of ML methodologies and critical appraisal of AI system limitations and ethical risks.
- **Establish Ethical AI Governance Frameworks:** Adopt formal governance structures for the evaluation, procurement, and ongoing monitoring of AI systems in library services, incorporating bias auditing, privacy impact assessment, and algorithmic transparency requirements.
- **Reposition Professional Labour:** Strategically redirect professional expertise toward functions that AI cannot replicate—complex reference mediation, collection curation, community engagement, ethical oversight, and critical evaluation of algorithmic outputs.
- **Engage Actively with Standards Development:** Participate in the development of AI-related standards for bibliographic data, metadata interoperability, and linked data vocabularies to ensure that LIS professional values are embedded in the technical infrastructure.
- **Pursue Interdisciplinary Research Collaborations:** Build research partnerships with computer science, cognitive science, science and technology studies, and critical data studies to produce rigorous, contextually grounded evaluations of AI applications in library and archival contexts.
- **Advocate for Equitable AI Access:** Engage in policy advocacy to ensure that AI-mediated information services do not exacerbate existing digital divides or concentrate knowledge access in ways that undermine the library's mission of universal intellectual access.

The research agenda for LIS scholarship must be equally ambitious. Priority areas include: the development of bias evaluation methodologies tailored to bibliographic

and archival data; longitudinal studies of user trust calibration in AI-mediated reference services; critical analysis of the epistemological implications of LLM-mediated information access for diverse user communities; investigation of the professional identity transformations associated with AI adoption; and the design of AI governance frameworks that reflect the distinctive values and institutional commitments of the LIS profession [59].

Conclusion

Artificial Intelligence and Machine Learning represent transformative forces that are reshaping every dimension of Library and Information Science practice and scholarship. From automated cataloguing to intelligent retrieval, from digital preservation to generative reference services, AI technologies are augmenting and, in some contexts, displacing human professional labour at a pace and scale that demands rigorous critical engagement from the LIS community. This chapter has argued that effective engagement with AI in LIS requires neither uncritical techno-enthusiasm nor reactionary resistance, but rather a disciplined critical praxis informed by the profession's foundational values: intellectual freedom, equitable access, privacy, and the stewardship of recorded knowledge as a public good. The epistemological sophistication that characterizes LIS scholarship—its attention to the social construction of knowledge, the politics of representation, and the institutional embedding of information systems—equips the profession uniquely to navigate the challenges and seize the opportunities of the AI moment. The most consequential question confronting LIS in the age of AI is not whether intelligent systems will transform the profession—that transformation is already underway—but rather who will govern that transformation, in whose interests, and according to what values. That question is, fundamentally, a question for library and information scientists.

Acknowledgement

We are very much thankful to the authors of different publications as many new ideas are abstracted from them. Authors also express gratefulness to their colleagues and family members for their continuous help, inspirations, encouragement, and sacrifices without which this work could not be executed. Finally, the main target of this work will not be achieved unless it is used by research institutions, students, research scholars, and authors in their future works. The authors will remain ever grateful to Dr. Neelu Singh, Director, ICFRE Tropical Forest Research Institute, Jabalpur, Principal, Jabalpur Engineering College, Jabalpur & Principal Government Science College, Jabalpur who helped by giving constructive suggestions for this work. The authors are also responsible for any possible errors and shortcomings, if any in the paper, despite the best attempt to make it immaculate.

References

1. Borgman, C. L. (2015). Big data, little data, no data: Scholarship in the networked world. MIT Press.
2. Svenonius, E. (2000). The intellectual foundation of information organization. MIT Press.
3. Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
4. Salton, G., & McGill, M. J. (1983). Introduction to modern information retrieval. McGraw-Hill.
5. Lindberg, D. A. B., & Humphreys, B. L. (1995). The unified medical language system and the challenge of rational access to biomedical information. *Yearbook of Medical Informatics*, 1(1), 78–89.
6. Smith, L. C. (1987). Artificial intelligence and information retrieval. *Annual Review of Information Science and Technology*, 22, 41–77.
7. Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22–29.
8. Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30(1–7), 107–117.
9. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
10. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171–4186).
11. Tait, E., & Pierson, C. M. (2022). Chatbots, AI and scholarly communication: Implications for academic libraries. *Journal of Academic Librarianship*, 48(3), 102513.
12. Golub, K., Hagelbäck, J., & Ardö, A. (2020). Automating subject classification of scholarly literature: Machine learning and subject heading assignment. *Journal of Data and Information Science*, 5(4), 19–47.
13. Suominen, O., Inkinen, J., & Lappalainen, M. (2022). Annif: DIY automated subject indexing using multiple algorithms. *LIBER Quarterly*, 29(1), 1–25.
14. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
15. Bianchi, F., Terragni, S., & Hovy, D. (2021). Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. In *Proceedings of ACL-IJCNLP 2021* (pp. 759–766).
16. Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT Press.

17. Eckert, K., & Pohl, A. (2022). Linking library data: Technical and policy challenges in federated authority file management. *Journal of Library Metadata*, 22(1–2), 1–28.
18. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
19. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of EMNLP-IJCNLP 2019* (pp. 3982–3992).
20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
21. Taylor, A. G., & Joudrey, D. N. (2018). *The organization of information* (4th ed.). Libraries Unlimited.
22. Tennant, R. (2002). MARC must die. *Library Journal*, 127(17), 26–28.
23. Suominen, O. (2019). Annif: DIY automated subject indexing using multiple algorithms. *Code4Lib Journal*, 46.
24. Mäkinen, S., Inkinen, J., & Suominen, O. (2022). Automated DDC classification of Finnish fiction using fine-tuned SciBERT. *Cataloging & Classification Quarterly*, 60(8), 759–781.
25. Kroeger, A. (2013). The road to BIBFRAME: The evolution of the idea of bibliographic transition into a post-MARC future. *Cataloging & Classification Quarterly*, 51(8), 873–890.
26. Pattuelli, M. C., & Hwang, K. (2021). *Linked data for libraries, archives and museums: How to clean, link and publish your metadata*. Facet Publishing.
27. Van Hooland, S., & Verborgh, R. (2014). *Linked data for libraries, archives and museums*. Facet Publishing.
28. Robertson, S. E., & Zaragoza, H. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4), 333–389.
29. Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense passage retrieval for open-domain question answering. In *Proceedings of EMNLP 2020* (pp. 6769–6781).
30. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.
31. Tripathi, M., & Shukla, A. (2023). Generative AI and library services: Opportunities, challenges, and emerging frameworks. *The Electronic Library*, 41(5), 598–614.

32. Desrosiers, C., & Karypis, G. (2011). A comprehensive survey of neighborhood-based recommendation methods. In F. Ricci et al. (Eds.), *Recommender systems handbook* (pp. 107–144). Springer.
33. Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
34. Coffman, S., & Arret, L. (2004). To chat or not to chat—taking another look at virtual reference. *Searcher*, 12(7), 38–46.
35. Dubicki, E. (2020). Chatbots in libraries: From novelty to valued service. *Reference Services Review*, 48(1), 55–72.
36. Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI-powered dialogue tools impact scholarly reference services? *Journal of Academic Librarianship*, 49(1), 102631.
37. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38.
38. Lavoie, B. (2014). *The open archival information system (OAIS) reference model: Introductory guide* (2nd ed.). Digital Preservation Coalition.
39. Pennock, M., & McKinney, P. (2021). The digital preservation maturity model: Evaluating how well organisations manage digital preservation challenges. In *Proceedings of iPRES 2021*.
40. Breuel, T. M. (2017). High performance text recognition using a hybrid convolutional-LSTM implementation. In *Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition* (pp. 11–16).
41. Seguin, B., Costiner, L., diLenardo, I., & Kaplan, F. (2018). New means for visual exploration of archives. In *Proceedings of the Digital Humanities 2018 Conference*.
42. Freire, N., Isaac, A., & Robson, G. (2018). Technical sustainability of aggregation in Europeana. In *Proceedings of the IFLA World Library and Information Congress 2018*.
43. Kaur, M., & Kaur, N. (2021). Evidence-based librarianship and machine learning: Transforming collection development in academic libraries. *Collection Building*, 40(4), 158–166.
44. Knoth, P., & Zdrahal, Z. (2012). CORE: Three access levels to underpin open access. *D-Lib Magazine*, 18(11/12).
45. Berman, S. (1971). *Prejudices and antipathies: A tract on the LC subject heads concerning people*. Scarecrow Press.
46. Olson, H. A. (2002). *The power to name: Locating the limits of subject representation in libraries*. Springer.
47. Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

48. Bates, M. J. (2017). Information and knowledge: An evolutionary framework for information science. *Information Research*, 10(4).
49. Zimmer, M. (2013). Patron privacy in the 2.0 era: Principles and practice for a networked world. *Journal of Information Ethics*, 22(1), 11–29.
50. IFLA. (2015). IFLA statement on privacy in the library environment. International Federation of Library Associations and Institutions.
51. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). 'Why should I trust you?': Explaining the predictions of any classifier. In *Proceedings of KDD 2016* (pp. 1135–1144).
52. Huvila, I. (2023). The professionalism of information work in the age of automation. *Journal of Documentation*, 79(1), 17–31.
53. Budd, J. M. (2021). The education of information professionals in the age of AI: Competencies, challenges, and futures. *Journal of Education for Library and Information Science*, 62(3), 277–290.
54. OpenAI. (2023). GPT-4 technical report. arXiv preprint arXiv:2303.08774.
55. Bowman, S. R. (2022). Eight things to know about large language models. arXiv preprint arXiv:2304.00612.
56. Van Noorden, R. (2023). More than 200 scientists have used ChatGPT to help write papers: Why and how? *Nature*, 622, 234–237.
57. Association of College and Research Libraries. (2016). Framework for information literacy for higher education. American Library Association.
58. Flohr, R., & Rehm, G. (2023). Applying large language models to archival finding aid generation: A case study with German municipal archives. *Archival Science*, 23, 119–141.
59. Hjørland, B. (2018). Library and information science (LIS). In B. Hjørland & C. Gnoli (Eds.), *ISKO Encyclopedia of Knowledge Organization*. ISKO.