Article:

Pinar, A. orcid.org/0000-0002-1716-5114 and Cox, A. orcid.org/0000-0002-2587-245X
(2025) An analysis of artificial intelligence (AI) capability in libraries and archives. Cataloging & Classification Quarterly. ISSN: 0163-9374

https://doi.org/10.1080/01639374.2025.2539790

An analysis of Artificial Intelligence (AI) capability in libraries and archives

Abdulhalik Pinar[1]

Andrew Cox[2]

Abstract

This paper seeks to evaluate the AI capability of libraries and archives using a qualitative content analysis of 54 case studies of AI uses published between 2018 and 2024. It is framed by the model of AI capability proposed by Mikalef and Gupta[3]. The findings of the analysis largely confirm the model, but suggest that there are many gaps in library and archive AI capability, especially in areas such as infrastructure and technical resources, data issues arising from metadata inconsistencies, and financial resources.

Introduction

Rapid developments in artificial intelligence (AI) especially after the mass availability of generative AI from late 2022, have led to many predictions about AI's potential to have a significant impact on sectors from health to finance. A number of authors have written about the potential impact on archives and libraries[4]. Some of these proposed impacts relate to new roles for librarians such as in AI literacy training. But impacts also include the application of AI to library services such as for metadata creation or other aspects of collection management and content discovery. Of course, while there are many potential benefits to using AI, there are also significant barriers, including costs, skills shortages, and ethical concerns surrounding this controversial technology.

In the absence of comprehensive studies of the sector wide adoption of AI in libraries and archives, perhaps our best insight into the realities of AI adoption are the many individual published case studies in which authors describe applications in a particular library or context. These case studies may not offer a comprehensive picture of the adoption of AI, but they do give detail of what specific tools are in use and for what purposes they are being used. They also give in-depth insights into the considerable challenges faced when adopting these technologies. Mannheimer et al.[5] carried out a wide ranging review of the same sort of data to explore how AI is being used in libraries and archives,

---

[1] Dr Abdulhalik Pinar, Faculty of Political Economy, Department of Management Information Systems, Harran University, Turkey. He is a TÜBİTAK-funded visiting scholar at University of Sheffield, supported under the International Postdoctoral Research Fellowship Programme (2219). ORCID: 0000-0002-1716-5114 Email: abdulhalik.pinar@harran.edu.tr.

[2] Dr Andrew M. Cox, Information School, University of Sheffield, UK. ORCID: 0000-0002-2587-245X. Email: a.m.cox@sheffield.ac.uk

[3] Patrick Mikalef and Manjul Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance', *Information & Management* 58, no. 3 (2021): 103434.

[4] Ryan Cordell, 'Machine Learning + Libraries', n.d.; Andrew M Cox and Suvodeep Mazumdar, 'Defining Artificial Intelligence for Librarians', *Journal of Librarianship and Information Science* 56, no. 2 (2024): 330–40; Sander Münster et al., 'Artificial Intelligence for Digital Heritage Innovation: Setting up a R&d Agenda for Europe', *Heritage* 7, no. 2 (2024): 794–816; Gaurav Shinde et al., 'AI in Archival Science--A Systematic Review', *arXiv Preprint arXiv:2410.09086*, 2024, https://doi.org/10.48550/arXiv.2410.09086.

[5] Sara Mannheimer et al., 'Responsible AI Practice in Libraries and Archives: A Review of the Literature', *Information Technology and Libraries* 43, no. 3 (2024).

especially in large academic settings. They found that most projects focus on digitized and born digital text and images, mainly for tasks like metadata creation and reference support. However, they also observed that ethical issues such as privacy, consent, social justice, and accessibility are often overlooked, with most discussions focused narrowly on technical performance. The current paper extends their work, with a particular interest in the acknowledged challenges libraries have faced in AI adoption.

While not all of the case studies focus we analyse relate directly to cataloguing or classification, the central concern of this journal, many contribute to the capabilities that make AI-supported cataloguing possible. Tasks such as OCR, entity recognition, and metadata extraction form upstream inputs for subject indexing and descriptive cataloguing, especially for digital collections. Other cases illustrate broader organisational capacities such as cross-departmental collaboration, data infrastructure, and strategic alignment that are essential for implementing AI in cataloguing workflows. Cataloguing does not operate in isolation; rather, it depends on an ecosystem of tools, resources, and institutional capability. Looking at how AI is used in related areas helps us better understand what makes AI-supported cataloguing and metadata creation possible in practice.

As an analytical lens to ground the study, this paper adopts the AI capability model developed by Mikalef and Gupta[6]. This is because the model offers a theoretically robust, comprehensive framework that has been widely used across different sectors. This model enables a multidimensional analysis of institutional capacity related to AI applications in knowledge organizations. In an assessment of the role of AI technologies in libraries and archives, factors such as data infrastructure, lack of necessary technical skills, and organizational transformation capacity were identified as the main structural needs associated with AI in such institutions[7]. In this context, the framework provided by the Mikalef and Gupta model appears to align with the case studies examined in this study. Other models developed to evaluate AI capability also exist; however, most of them are designed for specific sectors. For example, in the model developed by Chowdhury et al.[8] for the field of human resource management, AI capability is addressed through sector-specific organizational goals such as employee adoption, workforce transformation, governance, and ethics. The fact that this framework is shaped by the internal dynamics of a particular sector limits its direct applicability to different organizational structures. In contrast, the Mikalef and Gupta model focuses on more general organizational resources such as data infrastructure, technical and human skills, and organizational transformation capacity without being confined to a specific sectoral context, making it more easily adaptable to a variety of settings, including knowledge organisations.

Alternatively, AI maturity models and readiness models were also considered, but they were found to be of quite limited use given the aim of this study. For example, the maturity model developed by Schumacher, Erol, and Sihn[9] aims to position institutions along a continuum of stages from initial to advanced levels of AI implementation. The applicability of such models requires a broad dataset that includes institutions at different stages of development; however, the cases examined in this study consist of institutions that have already implemented AI. Similarly, readiness models are designed for evaluating institutions that have not yet adopted AI, and are therefore not suitable for analysing

[6] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.
[7] Andrew Cox, 'The Implications for Use of Artificial Intelligence in Libraries and Education: An Introductory Overview', in *New Horizons in Artificial Intelligence in Libraries* (Walter de Gruyter GmbH & Co KG, 2024).
[8] Soumyadeb Chowdhury et al., 'Unlocking the Value of Artificial Intelligence in Human Resource Management through AI Capability Framework', *Human Resource Management Review* 33, no. 1 (2023): 100899.
[9] Andreas Schumacher, Selim Erol, and Wilfried Sihn, 'A Maturity Model for Assessing Industry 4.0 Readiness and Maturity of Manufacturing Enterprises', *Procedia Cirp* 52 (2016): 161–66.

institutions already engaged in active use. In a review conducted within the field of library and information science, 12 different maturity models were analysed; however, the vast majority of these models were found to focus on areas such as digital preservation, digital libraries, or research data management, and are not directly related to AI capability[10]. For these reasons, the Mikalef and Gupta model offers a more appropriate framework for the aim of evaluating institutional capacity in this study.

This model is useful because it proposes a generalised framework of the resources needed for effective AI implementation, dividing them into three main categories: tangible resources, human skills and intangible resources. The model can help us think systematically about the resources library and archive case study authors refer to. Validating, or extending the model for the library and archive context, would be a useful contribution for future research and practice in our field. Based on this capability model, our study aimed, therefore, to answer the following four research questions:

1. What uses of AI are being made by libraries and archives?
2. What issues and barriers to the effective use of AI are discussed?
3. Do libraries have AI capability and in what areas do they lack the necessary resources?
4. Does the AI capability model of Mikalef and Gupta[11] predict the key aspects of capability in a library and archive context or are there other types of resource needed?

To answer these research questions, a qualitative content analysis was conducted of 54 case studies of AI applications published between 2018 and 2024, drawn from peer-reviewed journals, academic book chapters, conference proceedings, and some grey literature. Each case study was systematically reviewed to identify the type of institution, AI tools used, purposes, and challenges encountered.

Literature review

While the literature lacks comprehensive surveys of AI use in the library and archive sector internationally, there have been several surveys of librarians' readiness and attitudes to AI which may give us some insights into library and archive capabilities. Survey results suggest that cataloguing and metadata creation represent primary areas for AI application within libraries. Huang[12], for example, drawing on responses from 472 academic librarians in Taiwan, identified classification and subject indexing as key tasks for AI use. Lund et al.[13]'s survey data also suggests that applications in search and cataloguing are likely. Its use in operational efficiency and workflow automation is often mentioned too[14]. Library chatbots have also been seen as promising areas of development in studies[15].

AI is often portrayed as having the potential to significantly enhance library services, yet its adoption is often hindered by limited technical knowledge among librarians and inadequate technological infrastructure. These barriers are consistently identified across a range of studies from different

[10] Amit Tiwari and Devika P Madalli, 'Maturity Models in LIS Study and Practice', *Library & Information Science Research* 43, no. 1 (2021): 101069.
[11] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.
[12] Yuan-Ho Huang, 'Exploring the Implementation of Artificial Intelligence Applications among Academic Libraries in Taiwan', *Library Hi Tech* 42, no. 3 (2024): 885–905.
[13] Brady D Lund et al., 'Perceptions toward Artificial Intelligence among Academic Library Employees and Alignment with the Diffusion of Innovations' Adopter Categories', *College & Research Libraries* 81, no. 5 (2020): 865.
[14] Leticia Antunes Nogueira et al., 'Artificial Intelligence, Real Library' (NTNU, 2024).
[15] Cong Xu and Sandie Loo, 'A Review of Artificial Intelligence Applications in Libraries in Southeast Asia: Where Are We Now?', *Reference Services Review*, 2025, https://doi.org/10.1108/RSR-06-2024-0027.

regions. A common finding is that many librarians feel unprepared to work with AI technologies or resolve related issues, which highlights a pressing need for professional development and institutional support[16]. Even in well resourced research libraries, uncertainty and hesitation persist due to unfamiliarity with AI tools and concerns around responsible implementation[17]. In Global South countries, these challenges are compounded by structural limitations; for example, outdated systems, poor internet connectivity, and lack of funding are recurrent issues[18]. Regional studies further emphasize that infrastructure-related barriers are particularly severe in parts of Southeast Asia, where underdeveloped systems and bureaucratic obstacles often delay progress[19].

Financial limitations also present significant challenges for libraries trying to adopt AI according to surveys of librarians. Many libraries find the initial costs for AI technologies including specialized equipment, software licenses, and ongoing maintenance too high, especially given tight budgets[20]. This problem is even more pronounced in libraries from Global South countries, where limited financial resources restrict their ability to invest in necessary infrastructure or staff training for effective AI adoption[21].

Another area of challenge is the difficulty of managing data when adopting AI technologies. Issues such as inconsistent metadata, lack of standardized data structures, and limited interoperability between systems complicate the effective use of AI tools and increase the workload for library staff[22]. Additionally, inaccuracies from AI-driven digitization, especially in texts extracted from older or damaged documents, often require significant manual corrections, reducing the benefits of automation[23]. Another common challenge is the lack of standardized vocabularies and data structures, which prevents AI systems from working efficiently and reliably, further increasing the workload on librarians[24].

Ethical issues also strongly influence libraries' decisions about adopting AI. Librarians often express worries about protecting users' privacy, especially when AI tools handle sensitive personal data or

[16] Leo S Lo, 'Evaluating AI Literacy in Academic Libraries: A Survey Study with a Focus on US Employees', 2024; Clarivate, 'Pulse of the Library 2024', 2024, http://10.0.55.242/pulse.of.the.library.2024.
[17] LS Lo and CH Vitale, 'Evolving AI Strategies in Libraries: Insights from Two Polls of ARL Member Representatives over Nine Months', 2024.
[18] Muhammad Yousuf Ali et al., 'Artificial Intelligence Application in University Libraries of Pakistan: SWOT Analysis and Implications', *Global Knowledge, Memory and Communication* 73, no. 1/2 (2024): 219–34; Huang, 'Exploring the Implementation of Artificial Intelligence Applications among Academic Libraries in Taiwan'.
[19] Xu and Loo, 'A Review of Artificial Intelligence Applications in Libraries in Southeast Asia: Where Are We Now?'; Ebiere Diana Orubebe, Emmanuel Adeniyi Oloniruha, and Bolaji David Oladokun, 'Adoption and Utilization of Artificial Intelligence in Academic Libraries: Challenges and Opportunities in Developed and Developing Nations', *International Journal of Knowledge Content Development & Technology* 14, no. 3 (2024).
[20] Ali et al., 'Artificial Intelligence Application in University Libraries of Pakistan: SWOT Analysis and Implications'; Clarivate, 'Pulse of the Library 2024'; Huang, 'Exploring the Implementation of Artificial Intelligence Applications among Academic Libraries in Taiwan'.
[21] Orubebe, Oloniruha, and Oladokun, 'Adoption and Utilization of Artificial Intelligence in Academic Libraries: Challenges and Opportunities in Developed and Developing Nations'; Xu and Loo, 'A Review of Artificial Intelligence Applications in Libraries in Southeast Asia: Where Are We Now?'
[22] Ali et al., 'Artificial Intelligence Application in University Libraries of Pakistan: SWOT Analysis and Implications'; Huang, 'Exploring the Implementation of Artificial Intelligence Applications among Academic Libraries in Taiwan'; Xu and Loo, 'A Review of Artificial Intelligence Applications in Libraries in Southeast Asia: Where Are We Now?'
[23] Caroline Saccucci and Abigail Potter, '16 Assessing Machine Learning for Cataloging at the Library of Congress', in *New Horizons in Artificial Intelligence in Libraries*, ed. Edmund Balnaves et al. (De Gruyter Saur, n.d.), 227–38, https://doi.org/10.1515/9783111336435-017.
[24] Nogueira et al., 'Artificial Intelligence, Real Library'.

track user behaviour[25]. There is also widespread concern about potential biases in AI algorithms, which could lead to misinformation or unfair outcomes, causing hesitation in fully trusting AI-generated information[26].

Human factors, particularly resistance from library staff, significantly affect AI adoption. Librarians often express scepticism or anxiety toward adopting new technologies, fearing changes to their daily work routines or possible replacement of their roles by AI tools[27]. This resistance is often linked to uncertainty about AI's reliability, concerns over increased workloads during initial implementation, and general discomfort with rapidly evolving technologies[28]. Additionally, librarians frequently express anxiety about the risk of job losses and changes to traditional roles, particularly in cataloguing and reference work, as AI tools become more prevalent[29].

Libraries also appear to face challenges related to strategic direction and institutional support when adopting AI. Librarians frequently report confusion or uncertainty due to a lack of clear guidelines, strategies, or support from library management[30]. Without well-defined strategies or institutional backing, librarians struggle to integrate AI effectively into their services, leading to partial or unsuccessful implementations[31].

AI Capability

Mikalef and Gupta[32]'s AI Capability Model is a framework identifying the resources needed for AI split across three types: tangible resources, human skills, and intangible resources. The authors identify three sets of tangible resources, namely: data, technology and basic resources. As regards data, they suggest that AI requires high-quality, structured, and unstructured data at appropriate granularity, integrated from internal and external sources. The model emphasizes data accessibility, cleansing, and secure storage. Technology resources consist of hardware (e.g., GPUs, scalable storage) and software (e.g., machine learning platforms, Natural language processing tools) to process data and run AI algorithms. Investments in parallel computing and cloud-based AI services (e.g., APIs for vision or speech recognition) are seen as critical. Basic Resources include financial resources and

---

[25] Clarivate, 'Pulse of the Library 2024'; Lo, 'Evaluating AI Literacy in Academic Libraries: A Survey Study with a Focus on US Employees'; Xu and Loo, 'A Review of Artificial Intelligence Applications in Libraries in Southeast Asia: Where Are We Now?'
[26] Ali et al., 'Artificial Intelligence Application in University Libraries of Pakistan: SWOT Analysis and Implications'; Huang, 'Exploring the Implementation of Artificial Intelligence Applications among Academic Libraries in Taiwan'; Clarivate, 'Pulse of the Library 2024'.
[27] Lo, 'Evaluating AI Literacy in Academic Libraries: A Survey Study with a Focus on US Employees'; Lo and Vitale, 'Evolving AI Strategies in Libraries: Insights from Two Polls of ARL Member Representatives over Nine Months'; Ali et al., 'Artificial Intelligence Application in University Libraries of Pakistan: SWOT Analysis and Implications'.
[28] Orubebe, Oloniruha, and Oladokun, 'Adoption and Utilization of Artificial Intelligence in Academic Libraries: Challenges and Opportunities in Developed and Developing Nations'; Huang, 'Exploring the Implementation of Artificial Intelligence Applications among Academic Libraries in Taiwan'.
[29] Ali et al., 'Artificial Intelligence Application in University Libraries of Pakistan: SWOT Analysis and Implications'; Lo, 'Evaluating AI Literacy in Academic Libraries: A Survey Study with a Focus on US Employees'; Clarivate, 'Pulse of the Library 2024'; Orubebe, Oloniruha, and Oladokun, 'Adoption and Utilization of Artificial Intelligence in Academic Libraries: Challenges and Opportunities in Developed and Developing Nations'.
[30] Lo, 'Evaluating AI Literacy in Academic Libraries: A Survey Study with a Focus on US Employees'; Huang, 'Exploring the Implementation of Artificial Intelligence Applications among Academic Libraries in Taiwan'.
[31] Orubebe, Oloniruha, and Oladokun, 'Adoption and Utilization of Artificial Intelligence in Academic Libraries: Challenges and Opportunities in Developed and Developing Nations'.
[32] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.

time to dedicate to AI projects, including funding for experimentation, staffing, and long-term development.

The second set of resources, human skills, are of two types, namely technical skills and business skills. Not surprisingly they consider technical skills as important: they imply proficiency in AI techniques (e.g., machine learning, deep learning) and data science (e.g., data analysis, algorithm development). But the authors also give emphasis to business skills: such as the leadership abilities to align AI with organizational goals, identify use cases, manage cross-functional teams, and communicate AI's value to stakeholders.

The final set of resources required for AI capability according to Mikalef and Gupta[33] are intangible resources. These are organizational traits that enable AI adoption. Again there are three components. Inter-departmental coordination is about aligning work across teams in AI projects to organizational priorities. Organizational change capacity is about having the agility to adapt workflows, policies, and cultures to integrate AI. Finally risk proclivity is the willingness to pursue high-risk, high-reward AI initiatives.

The model has an inherent plausibility, and it was decided to adopt it as a framework to reflect on the analysis. Conceptually, the model is grounded in the Resource-Based View (RBV) of the firm, which argues that organizations can gain a competitive advantage by effectively using resources that are valuable, rare, difficult to imitate, and non-substitutable[34]. Although RBV has been widely applied to understand how internal capabilities lead to superior performance, recent work has highlighted its limitations. For example, Zahra[35] points out that RBV often assumes stable access to resources and clear strategic rationality, but firms especially those in uncertain or resource constrained environments where organisations must constantly adapt, experiment, and exercise judgment in how they manage and recombine what they have. Freeman et al.[36] take this further, arguing that RBV tends to overlook the role of values, ethics, and stakeholder relationships in shaping what is seen as valuable. They suggest that long-term, trust-based stakeholder relationships are not just context for strategy but are strategic resources. Similarly, while Barney et al.[37] do not explicitly focus on ethical or normative dimensions, they do stress that firms rely on contributions from a wide range of stakeholders and that understanding value creation means looking beyond shareholders alone. Adding to this discussion, Moderno et al.[38] argue that AI and robotic process automation capabilities challenge the static assumptions of classical RBV, requiring organizations to integrate cultural alignment, organizational learning, and flexible resource recombination into their strategic models. These newer perspectives are especially relevant in public and memory institutions such as libraries and archives, where AI adoption is not just about efficiency or performance but also about alignment with institutional purpose, trustworthiness, and long term public value. Our use of the AI capability model builds on this expanded understanding of RBV, recognizing that in these settings, culture, transparency, and responsibility are not separate from strategy they are part of what enables it.

[33] Mikalef and Gupta.
[34] Jay Barney, Mike Wright, and David J Ketchen Jr, 'The Resource-Based View of the Firm: Ten Years after 1991', *Journal of Management* 27, no. 6 (2001): 625–41.
[35] Shaker A Zahra, 'The Resource-Based View, Resourcefulness, and Resource Management in Startup Firms: A Proposed Research Agenda', *Journal of Management* 47, no. 7 (2021): 1841–60.
[36] R Edward Freeman, Sergiy D Dmytriyev, and Robert A Phillips, 'Stakeholder Theory and the Resource-Based View of the Firm', *Journal of Management* 47, no. 7 (2021): 1757–70.
[37] Barney, Wright, and Ketchen Jr, 'The Resource-Based View of the Firm: Ten Years after 1991'.
[38] Osvaldo Braz dos Santos Moderno, Antonio Carlos Braz, and Paulo Tromboni de Souza Nascimento, 'Robotic Process Automation and Artificial Intelligence Capabilities Driving Digital Strategy: A Resource-Based View', *Business Process Management Journal* 30, no. 1 (2024): 105–34.

Method

This study is based on a qualitative content analysis applied to case studies of the use of artificial intelligence (AI) in libraries and archives. The case studies were identified through searches in databases including LISA (Library and Information Science Abstracts), LISTA (Library, Information Science & Technology Abstracts), Scopus, and Google Scholar. Searches were conducted using a combination of keywords: “artificial intelligence, AI, natural language processing” AND “libraries, archives, metadata, automated indexing”. Only case studies published in English were included, as the study aimed to capture internationally accessible literature.

To be included, the case studies needed to:

- Be presented in peer-reviewed journal articles, academic book chapters, conference proceedings, or institutional reports;
- Involve the use of AI tools or techniques in a library or archive context;
- Provide enough information about the purpose, application, and challenges of the AI use to shed light on our research questions.

A total of 54 case studies published between 2018 and 2024 were identified (See Figure 1 and appendix 1). There has been considerable debate about AI in the last few years and many conferences where case studies have been presented (such as the Fantastic Futures conference series) but it was decided to focus on work published as full papers where sufficient detail is given for planned analysis.

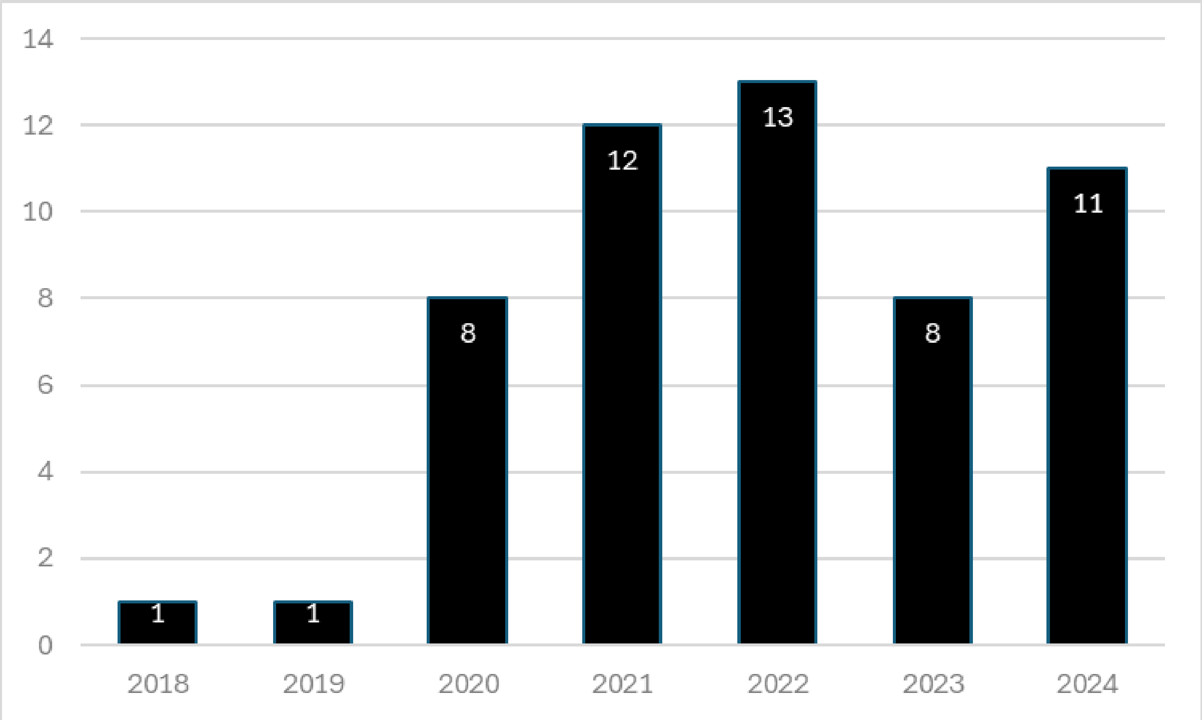

Figure 1: Distribution of case studies by years

Each case study was reviewed to understand how AI is being used across different types of institutions. We noted the type of library or archive, when the AI project was implemented, what kind of tools were used, and the main goals of the projects and the challenges that were mentioned.

The breakdown of institutions included in the study is as follows:

- Academic Libraries: 34
- National Libraries: 17

- Public Libraries: 2
- National Archives: 1

The validity of the findings of this study is limited due to the relatively small number of available published case studies. More fundamentally what is published is unlikely to represent the full range of activity. Only a minority of practitioners are motivated to publish their results in peer reviewed/ extended form. In addition, it is far more likely to publish results of complex projects that are deemed to be “innovative” than to describe practices that have become widely accepted. So the reader should not take the findings to represent the full range of what is happening, but they do offer in-depth insights into that can be usefully summarised in the context of our interest in confirming the AI capability model put forward by Mikalef and Gupta[39].

Findings

This review of 54 case studies published between 2018 and 2024 shows how libraries and archives have approached the use of artificial intelligence (AI) over time. In the earlier years (2018–2020), projects mainly focused on foundational technologies like Optical Character Recognition (OCR), basic machine learning models, and rule-based metadata tools. These were used to support digitization and to manage large volumes of text and scanned documents. Around 2021, there was a noticeable shift in interest in more advanced tools, especially transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT). These were applied to tasks such as improving metadata quality and handling multilingual content. However, this interest did not expand much further with only a few projects in 2022 through 2024 continued down that path. So, while some experimentation with these models occurred, they did not become a dominant trend. By 2024, publications focusing on foundational tools like OCR and basic Natural Language Processing (NLP) declined. This does not necessarily mean those tools were no longer in use, it may simply reflect a shift in what researchers chose to write about. Instead, the focus in 2024 appears to have turned toward broader applications, such as using AI to support user interaction, digital literacy, and integration into public-facing services. For instance, some projects explored large language models (LLMs) for classification, while others used AI in educational tools or chat based interfaces.

Most Used Artificial Intelligence Tools in Libraries and Archives

The rate of publication of papers about AI suggests that its adoption in libraries and archives is expanding, with many different applications ranging from text processing and visual data analysis to automatic classification and user interaction. The following overview highlights the AI tools most frequently used across analysed case studies. Figure 2 summarizes the frequency with which different categories of AI techniques, such as natural language processing, machine learning, and speech and vision, were referenced across the 54 case studies. These categories were derived from a manual review of each project, capturing whether a specific technique was substantively applied.

Text Processing and Metadata Enrichment is one important type of tool. Transformer-based models such as BERT and its derivatives (9 cases) are increasingly utilized for text understanding, metadata enrichment, and classification. For instance, the University of Illinois applied BERT to improve quality control for historical text digitization, enhancing the accuracy of metadata tagging[40]. Authors

[39] Mikalef and Gupta, ‘Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance’.
[40] Mariana Dias and Carla Teixeira Lopes, ‘Optimization of Image Processing Algorithms for Character Recognition in Cultural Typewritten Documents’, *ACM Journal on Computing and Cultural Heritage* 16, no. 4 (2023): 1–25.

from the National Library of Sweden describe the use of a customized KB-BERT model for multilingual metadata consistency[41].

Optical Character Recognition technologies (8 cases) play a critical role in digitizing historical documents and extracting textual content. Notably, the National Library of Finland implemented Tesseract OCR for 19th-century manuscripts, significantly improving character recognition accuracy[42]. The HathiTrust Digital Library similarly describes the employment of OCR for multilingual document processing, enhancing resource accessibility[43]. Natural Language Processing (NLP) (7 cases) facilitates automated metadata generation and text classification, as exemplified by the German National Library's optimized metadata workflows through NLP-based models, significantly increasing cataloguing efficiency[44].

Computer vision technologies (6 cases) support image-based metadata enrichment and cataloguing. Stanford University utilizes the Faster R-CNN model for automated cataloguing of artworks, significantly improving discoverability for researchers[45].

General Machine Learning techniques (6 cases) assist in classifying and managing large datasets. For example, the National Library of Norway has effectively automated its Dewey Decimal Classification workflows using machine learning algorithms[46].

Tesseract OCR (4 cases) is specifically effective in digitizing handwritten documents. Carnegie Mellon University Libraries enhanced archival workflows by integrating Tesseract OCR to digitize handwritten manuscripts[47].

AI-driven chatbots (4 cases) have become integral for user interaction and information retrieval. The University of Delaware Libraries, for instance, uses AI chatbots to streamline resource access and enhance user engagement[48].

Several specialized AI tools have also been implemented for specific archival needs. Annif (3 cases) is widely used for automatic subject indexing, and Transkribus (2 cases) demonstrates effectiveness in handwriting recognition[49]. Additionally, the Kratt tool (1 case) has been adopted by the National Library of Estonia for AI-driven metadata enrichment[50].

---

[41] Elisabeth Modden, 'Artificial Intelligence, Machine Learning and Bibliographic Control: DDC Short Numbers: Towards Machine-Based Classifying', *JLIS: Italian Journal of Library, Archives and Information Science= Rivista Italiana Di Biblioteconomia, Archivistica e Scienza Dell'informazione: 13, 1, 2022*, 2022, 256–64.
[42] Kimmo Kettunen, Mika Koistinen, and Jukka Kervinen, 'Ground Truth OCR Sample Data of Finnish Historical Newspapers and Journals in Data Improvement Validation of a Re-OCRing Process', *LIBER Quarterly: The Journal of the Association of European Research Libraries* 30, no. 1 (2020): 1–20.
[43] Dias and Lopes, 'Optimization of Image Processing Algorithms for Character Recognition in Cultural Typewritten Documents'.
[44] Modden, 'Artificial Intelligence, Machine Learning and Bibliographic Control: DDC Short Numbers: Towards Machine-Based Classifying'.
[45] Catherine Nicole Coleman, 'Computer Vision and Cultural Heritage Case Study 2' (Stanford: AEOLIAN, 2021), https://www.aeolian-network.net/wp-content/uploads/2022/04/Aeolian-Case-Study-2-April-2022.pdf.
[46] Modden, 'Artificial Intelligence, Machine Learning and Bibliographic Control: DDC Short Numbers: Towards Machine-Based Classifying'.
[47] Matthew Lincoln et al., 'CAMPI: Computer-Aided Metadata Generation for Photo Archives Initiative', 2020.
[48] Michelle Ehrenpreis and John DeLooper, 'Implementing a Chatbot on a Library Website', *Journal of Web Librarianship* 16, no. 2 (2022): 120–42.
[49] Mustak Ahmed, Mondrita Mukhopadhyay, and Parthasarathi Mukhopadhyay, 'Automated Knowledge Organisation: AI/ML-Based Subject Indexing System for Libraries.', *DESIDOC Journal of Library & Information Technology* 43, no. 1 (2023).
[50] Marit Asula et al., 'Kratt: Developing an Automatic Subject Indexing Tool for the National Library of Estonia', *Cataloging & Classification Quarterly* 59, no. 8 (2021): 775–93.

Large Language Models (LLMs) (1 case) are emerging as tools for dynamic metadata generation, with institutions like the German National Library exploring their potential in automating classification and enhancing retrieval processes[51].

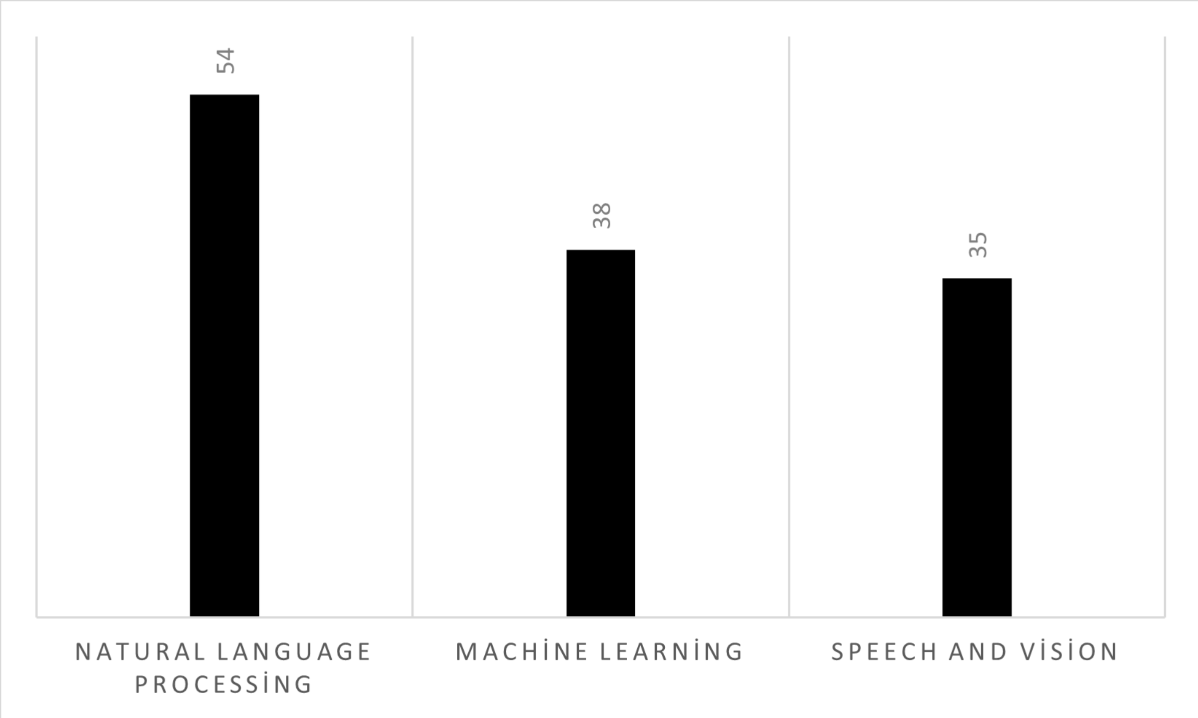

Figure 2: AI Tools in use

For analysis we used the AI family tree model created by Wheatley and Hervieux[52] as a guiding framework for organizing the different types of AI technologies we encountered. Originally developed for their library workshops on AI literacy, it groups tools like natural language processing, machine learning, and speech and vision into clear, visual clusters. In our study, we observed that Natural Language Processing (NLP) was the most frequently utilized AI category, appearing in 54 cases across the reviewed projects. This reflects the widespread adoption of language-focused tools such as BERT, chatbots, and subject indexing systems like Annif and Kratt technologies often used to support metadata enrichment, automated classification, and user interaction. Machine Learning (ML) followed with 38 instances, underpinning many foundational and functional components of library AI systems, including classification algorithms and the architecture of large language models. We also grouped tools related to Speech and Vision such as OCR (e.g., Tesseract), handwriting recognition (e.g., Transkribus), and speech-to-text systems under a unified category, which appeared in 35 cases.

Purposes of use

Our second analysis was of the purposes to which AI was being put. Again we endeavoured to place each case study in a single category. A significant proportion of the case studies prioritized Metadata Creation, Enrichment, and Indexing (15 cases, 28%). AI technologies, particularly transformer-based

---

[51] Modden, 'Artificial Intelligence, Machine Learning and Bibliographic Control: DDC Short Numbers: Towards Machine-Based Classifying'.

[52] Amanda Wheatley and Sandy Hervieux, 'Separating Artificial Intelligence from Science Fiction: Creating an Academic Library Workshop Series on AI Literacy', in *The Rise of AI: Implications and Applications for AI in Academic Libraries*, 2022, https://escholarship.mcgill.ca/concern/books/0r9678471.

NLP models like BERT, were commonly deployed to automate and enhance metadata quality. For instance, the National Library of Estonia successfully automated subject indexing and metadata generation for historical records through their AI-powered tool, Kratt, significantly improving operational efficiency[53].

An equally significant objective was digitization and operational automation (15 cases, 28%), primarily leveraging optical character recognition (OCR). Tools like Tesseract OCR have been widely employed, although challenges related to OCR quality and integration with existing legacy systems frequently surfaced[54].

Increasing Access to Collections (11 cases, 20%) was also a key goal, facilitated by AI-driven tools such as chatbots and computer vision models. While chatbots were referenced in isolated examples, they did not represent a widespread or analytically prominent trend across the dataset. For example, Hangzhou Public Library implemented AI chatbots to streamline user interaction and support, significantly easing demands on traditional reference services[55]. Similarly, the Frick Art Reference Library utilized computer vision technology to enhance discoverability within digitized collections, making resources more accessible to researchers[56].

Several institutions utilized AI specifically for collection analysis and to enhance research support capabilities (10 cases, 19%). Carnegie Mellon University Libraries' CAMPI —an AI-driven tool that supports metadata generation using computer vision— initiative is notable, employing computer vision to cluster and analyse visual archives, which significantly streamlined research processes[57].

A smaller subset of case studies focused specifically on gaining deeper insights from datasets, highlighting AI's potential to analyse patterns, trends, and relationships within archival materials, although these cases were fewer and indicated a niche but growing interest (3 cases, 5%).

Overall, these findings suggest a clear trend toward more sophisticated and ethically informed AI applications, moving beyond foundational tasks to strategic initiatives that align closely with broader institutional goals and user needs.

Challenges

The analysis of 54 case studies reveals a range of recurring challenges in applying AI within library and archive environments. Unlike the previous two analyses the case studies usually mentioned multiple aspects, so as many categories as appeared were included, rather than picking a single category. As shown in Figure 3, the most frequently mentioned were financial, data management, and ethical challenges, followed by technical, integration, metadata-related, and resource management issues.

[53] Asula et al., 'Kratt: Developing an Automatic Subject Indexing Tool for the National Library of Estonia'.
[54] Kettunen, Koistinen, and Kervinen, 'Ground Truth OCR Sample Data of Finnish Historical Newspapers and Journals in Data Improvement Validation of a Re-OCRing Process'; Dias and Lopes, 'Optimization of Image Processing Algorithms for Character Recognition in Cultural Typewritten Documents'.
[55] Bing Nie et al., 'How Does Ai Make Libraries Smart?: A Case Study of Hangzhou Public Library', in *Technological Advancements in Library Service Innovation* (IGI Global Scientific Publishing, 2022), 43–58.
[56] Ellen Prokop et al., 'AI and the Digitized Photoarchive: Promoting Access and Discoverability', *Art Documentation: Journal of the Art Libraries Society of North America* 40, no. 1 (2021): 1–20.
[57] Matthew Lincoln et al., 'CAMPI: Computer-Aided Metadata Generation for Photo Archives Initiative', 2020.

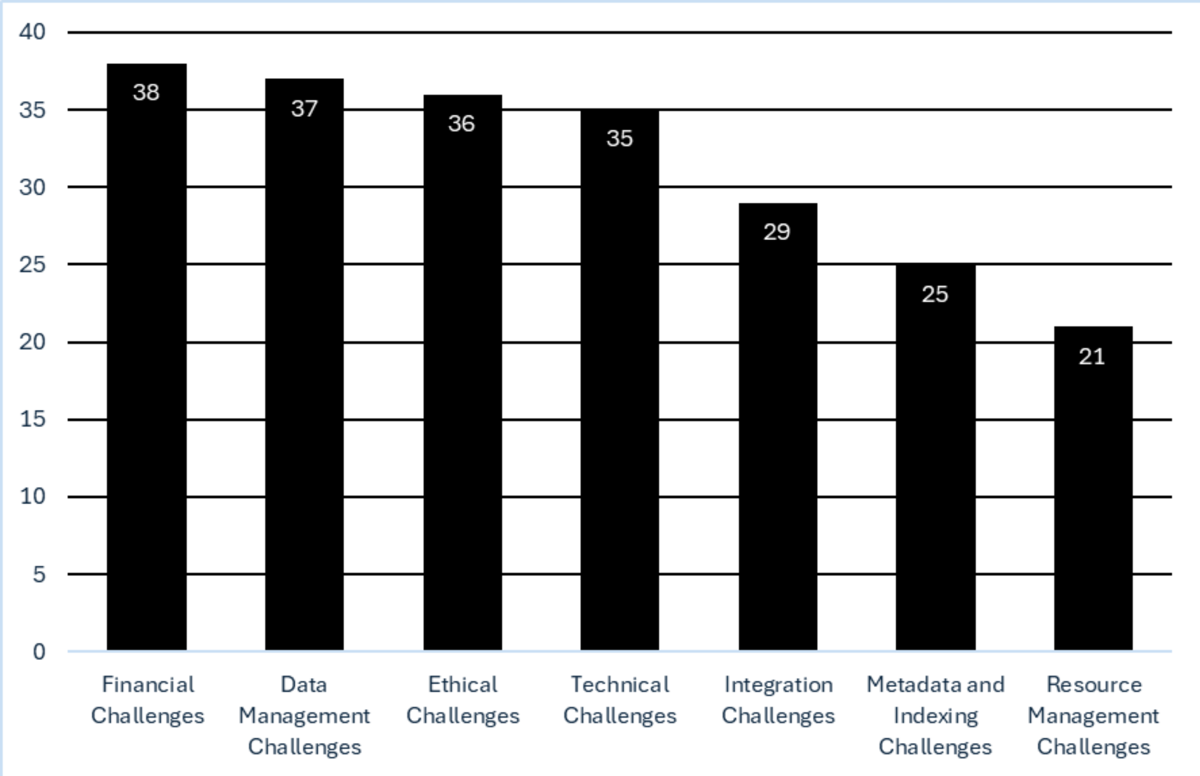

Figure 3: Frequency of challenges mentioned

Financial Challenges (38 out of 54 cases, 70%)

It seems from the case studies that financial constraints represent a significant barrier to successfully implementing and sustaining AI solutions in libraries and archives. The main financial concerns identified were high initial investment costs, unpredictable ongoing maintenance expenses, and limited institutional funding, which collectively threaten the sustainability and long-term viability of AI initiatives. High initial investment costs frequently prevent institutions from initiating or fully implementing AI projects. For example, Stanford University's Global Currents project faced substantial computational bottlenecks due to inadequate GPU (Graphics Processing Unit) resources, requiring significant upfront investment in specialized hardware and computing infrastructure to resolve[58].

Unpredictable ongoing maintenance expenses also create substantial obstacles, often causing projects to fail after initial success. San Jose State University Library suspended its AI-enabled chatbot despite promising early results, primarily due to mounting and unanticipated operational costs associated with the continued maintenance and upgrades[59].

Even when initial investment and ongoing maintenance are feasible, limited institutional funding structures create financial uncertainties. The National Library of Finland initially adopted the proprietary ABBYY FineReader for digitizing historical documents but later shifted to open-source Tesseract due to licensing expenses. However, this switch introduced significant hidden costs,

---

[58] Coleman, 'Computer Vision and Cultural Heritage Case Study 2'.
[59] Ehrenpreis and DeLooper, 'Implementing a Chatbot on a Library Website'.

particularly manual post-correction efforts, underscoring that even "free" AI solutions require adequate, sustainable financial support[60].

Financial sustainability remains a persistent challenge even for relatively well-funded institutions. The National Archives (UK) emphasized difficulties maintaining digital infrastructure capable of supporting large-scale AI-driven archival processing, despite receiving substantial public funding[61]. Similarly, institutions involved in the AEOLIAN project noted that externally funded pilot projects required robust long-term financial strategies to transition effectively into sustainable operational workflows[62].

Finally, it is important to note that it seemed that financial constraints disproportionately impact institutions in the Global South, highlighting and deepening the digital divide. For example, the University of Kalyani faced considerable difficulties securing sufficient funds to launch and sustain AI-based subject indexing systems and encountered significant problems retaining skilled personnel due to limited ability to offer competitive salaries and infrastructure support[63].

These examples collectively demonstrate how financial barriers—ranging from high upfront costs and ongoing maintenance unpredictability to restricted institutional funding—significantly affect the implementation, operational continuity, and scalability of AI projects across libraries and archives.

Data Management Challenges (37 out of 54 cases, 69%)

Data management emerged as one of the most prevalent and multifaceted challenges in AI implementation across libraries and archives. Key issues frequently identified in the papers include metadata inconsistencies, data sparsity, heterogeneity of datasets, and the varying quality of OCR outputs. These problems significantly affect machine learning model training, integration of data, and the usability of digital archives.

Metadata inconsistencies were a prominent and recurring challenge. The EyCon project, focused on early conflict photography and visual AI, reported significant issues with outdated metadata, inconsistent archival practices across institutions, and insufficient contextual information, particularly problematic in collections of sensitive historical content[64]. Similarly, the Frick Art Reference Library encountered substantial barriers due to sparse and semantically outdated metadata, which impeded the effectiveness of automated classification processes, necessitating extensive manual curation to avoid misclassifications[65].

Challenges related to metadata standardization and enrichment at scale further limited AI effectiveness. The CAMPI initiative (Computer-Aided Metadata Generation for Photo Archives) struggled specifically with the integration of AI-driven object detection into archival metadata workflows. They identified that inconsistent metadata structures across collections significantly

[60] Kettunen, Koistinen, and Kervinen, 'Ground Truth OCR Sample Data of Finnish Historical Newspapers and Journals in Data Improvement Validation of a Re-OCRing Process'.
[61] Lise Jaillant, Katherine Aske, and Annalina Caputo, 'The National Archives (UK): Case Study' (AEOLIAN, 2021).
[62] Tim Hutchinson, 'Natural Language Processing and Machine Learning as Practical Toolsets for Archival Processing', *Records Management Journal* 30, no. 2 (2020): 155–74.
[63] Ahmed, Mukhopadhyay, and Mukhopadhyay, 'Automated Knowledge Organisation: AI/ML-Based Subject Indexing System for Libraries.'
[64] Katherine Aske and Marina Giardinetti, '(Mis) Matching Metadata: Improving Accessibility in Digital Visual Archives through the EyCon Project', *ACM Journal on Computing and Cultural Heritage* 16, no. 4 (2023): 1–20.
[65] Koraljka Golub, 'Automated Subject Indexing: An Overview', *Cataloging & Classification Quarterly* 59, no. 8 (2021): 702–19.

constrained the automation success, highlighting a broader issue of lacking uniform standards across archival datasets[66].

OCR output quality posed an ongoing obstacle, limiting data usability for AI training and application. The National Library of Finland emphasized that OCR-generated errors introduced substantial noise into datasets, particularly affecting historical documents in varying conditions. These inaccuracies complicated downstream AI tasks and reduced overall reliability[67]. HathiTrust Digital Library faced similar difficulties, noting that inconsistent OCR quality across document types severely hampered character recognition accuracy and the effectiveness of text analysis[68].

Data sparsity was also frequently cited as a significant limitation. Stanford University's Global Currents project reported challenges with training computer vision models due to limited and uneven datasets, resulting in poor performance and increased computational requirements[69]. These constraints not only demand high-quality data but also entail substantial additional manual labelling and dataset correction efforts.

The heterogeneity of datasets further complicated effective data management and AI implementation. The University of Kalyani encountered substantial difficulties in managing and standardizing diverse bibliographic records when training a semi-automated subject indexing system. Inconsistent metadata structures across their records adversely affected indexing accuracy and performance, highlighting how data heterogeneity compounds challenges in resource-limited contexts[70]. Similarly, the AEOLIAN project noted that variations in archival record organization, absence of standardized vocabularies, and a lack of consistently context-rich metadata undermined the accuracy and integration of AI-based classification tools[71].

Collectively, these examples demonstrate how metadata inconsistencies, limited standardization, OCR quality issues, data sparsity, and heterogeneity of datasets significantly hinder AI adoption and operational success within library and archive environments.

Ethical Challenges (36 out of 54 cases, 67%)

As libraries and archives increasingly adopt artificial intelligence (AI), ethical concerns surrounding these technologies have become as significant as their technical and operational challenges. The key ethical issues identified in the analysis include algorithmic bias, lack of transparency, risks of decontextualization, and the limitations of current AI systems that necessitate extensive human intervention, restricting fully automated processes. These challenges become particularly critical when applying AI tools to sensitive or collections about historically marginalized communities.

Algorithmic Bias and Historical Inaccuracy

Algorithmic bias, particularly arising from historically skewed or incomplete datasets, is a major ethical concern. The Library of Congress's Newspaper Navigator project, for instance, highlighted those inaccuracies in OCR and biases within training data led to the misclassification and distortion of historical content, negatively affecting representation of marginalized or underrepresented

66 Jeremiah Flannery, 'Using NLP to Generate MARC Summary Fields for Notre Dame's Catholic Pamphlets', *International Journal of Librarianship* 5, no. 1 (2020): 20–35.
67 Kettunen, Koistinen, and Kervinen, 'Ground Truth OCR Sample Data of Finnish Historical Newspapers and Journals in Data Improvement Validation of a Re-OCRing Process'.
68 Allen Kim et al., 'Cleaning Dirty Books: Post-OCR Processing for Previously Scanned Texts', *arXiv Preprint arXiv:2110.11934*, 2021.
69 Coleman, 'Computer Vision and Cultural Heritage Case Study 2'.
70 Ahmed, Mukhopadhyay, and Mukhopadhyay, 'Automated Knowledge Organisation: AI/ML-Based Subject Indexing System for Libraries.'
71 Jaillant, Aske, and Caputo, 'The National Archives (UK): Case Study'.

communities[72]. Similarly, the Smithsonian Institution reported challenges related to culturally biased metadata generated by AI. When applied to global cultural collections, these biases reinforced Western-centric perspectives and led to misleading interpretations or mislabelling of non-Western materials[73].

Limitations of AI and Need for Extensive Human Intervention

Current AI systems frequently lack the capability to accurately interpret complex semantic contexts without extensive human oversight, significantly limiting their autonomous operation. The University of Ottawa emphasized the necessity of integrating domain expertise into AI workflows, noting that AI-generated metadata often suffered from semantic errors and contextual misunderstandings unless closely supervised by human experts[74]. A project at Carnegie Mellon University Libraries also demonstrated these limitations, revealing that AI-driven metadata generation tools consistently failed to fully capture document context, particularly in historical collections. The system frequently produced overly generic or misleading labels, requiring extensive human correction and validation[75].

Decontextualization Risks in AI-Driven Digitization

Another significant ethical concern involves the potential decontextualization of archival materials by AI systems, especially with documents related to sensitive historical events involving colonialism, violence, or trauma. The National Archives (UK) identified that their AI-assisted sensitivity review tools often failed to preserve contextual integrity, resulting in misrepresentations or overly simplified historical narratives[76]. The EyCon Project similarly reported significant challenges arising from applying generalized vocabularies to culturally specific imagery. The use of universalized terms led to classification errors, loss of original contextual meaning, and increased risk of marginalizing already underrepresented cultural perspectives[77].

Technical Challenges (35 out of 54 cases, 65%)

The use of artificial intelligence (AI) in libraries and archives faces substantial technical barriers, primarily related to computational demands, limitations of existing algorithms, OCR performance issues, and the complexity of configuring advanced AI tools. These technical difficulties often delay project implementation, negatively impact AI model accuracy, and significantly increase workload demands for technical and archival staff.

Computational resource limitations were frequently cited in the case studies. The Stanford Global Currents project, which applied computer vision techniques to medieval manuscripts, specifically highlighted insufficient GPU resources and data sparsity as significant technical constraints. These limitations directly impacted AI model performance, accuracy, and scalability[78]. Similarly, the University of Basel encountered considerable challenges optimizing classification algorithms due to

[72] Benjamin Lee, 'Compounded Mediation: A Data Archaeology of the Newspaper Navigator Dataset', *Digital Humanities Quarterly* 15, no. 4 (2021): 3055–62.
[73] Rebecca Dikow et al., 'Developing Responsible AI Practices at the Smithsonian Institution', *ARPHA Preprints* 4 (2023): e113335.
[74] Chris Oliver, 'Leveraging KOS to Extend Our Reach with Automated Processes', *Cataloging & Classification Quarterly* 59, no. 8 (2021): 868–74.
[75] Lincoln et al., 'CAMPI: Computer-Aided Metadata Generation for Photo Archives Initiative', 2020.
[76] Jaillant, Aske, and Caputo, 'The National Archives (UK): Case Study'.
[77] Aske and Giardinetti, '(Mis) Matching Metadata: Improving Accessibility in Digital Visual Archives through the EyCon Project'.
[78] Coleman, 'Computer Vision and Cultural Heritage Case Study 2'.

limited computational capacity, particularly when dealing with extensive, high-resolution image datasets[79].

The complexity of configuring and operating AI tools presents another prominent technical challenge. The University of Kalyani reported significant difficulty managing multiple backend algorithms for their automated subject indexing tool, Annif. The configuration required advanced technical skills, stable computing environments, and sophisticated data pipeline management, presenting substantial technical barriers in resource-constrained settings[80].

OCR technology limitations represent another significant technical barrier, particularly for degraded or non-standard historical documents. The National Library of Finland transitioned from proprietary OCR software to open-source Tesseract primarily due to budget constraints, yet they continued encountering substantial technical difficulties related to accuracy, dialect variations, and formatting inconsistencies in historical texts[81]. Similarly, Dias and Lopes[82] found that despite using advanced preprocessing techniques (such as adaptive thresholding and morphological filtering), OCR output quality significantly varied depending on document condition and type. Their research emphasized that enhancing OCR accuracy requires not only better software selection but also extensive iterative preprocessing and parameter tuning processes demanding substantial technical expertise.

Algorithmic complexities and inconsistencies in performance also represent persistent technical obstacles. The Library of Congress found BERT-based models unreliable in consistently classifying complex historical documents, particularly in tasks requiring nuanced contextual understanding. This necessitated ongoing manual correction and technical recalibration of models[83]. The University of Ottawa similarly reported technical difficulties when natural language processing (NLP) tools failed to align semantically with curated taxonomies, diminishing metadata accuracy and requiring frequent recalibration efforts[84].

Finally, these technical challenges become increasingly complicated when institutions deploy multiple AI technologies simultaneously—such as OCR, NLP, and image classification—within large-scale digitization projects. The compounded complexity significantly increases technical burdens, computational requirements, and the overall demands on institutional technical expertise and infrastructure.

Integration Challenges (29 out of 54 cases, 54%)

Integrating artificial intelligence (AI) into library and archival environments frequently encounters difficulties related to legacy infrastructures, rigid workflows, institutional adaptability, and complexities in embedding AI into public-facing services. In the case studies, many institutions find that while AI tools promise transformative benefits, effectively integrating these tools into existing operational systems remains complex and resource-intensive.

---

[79] Murielle Cornut, Julien Antoine Raemy, and Florian Spiess, 'Annotations as Knowledge Practices in Image Archives: Application of Linked Open Usable Data and Machine Learning', *ACM Journal on Computing and Cultural Heritage* 16, no. 4 (2023): 1–19.
[80] Ahmed, Mukhopadhyay, and Mukhopadhyay, 'Automated Knowledge Organisation: AI/ML-Based Subject Indexing System for Libraries.'
[81] Kettunen, Koistinen, and Kervinen, 'Ground Truth OCR Sample Data of Finnish Historical Newspapers and Journals in Data Improvement Validation of a Re-OCRing Process'.
[82] Dias and Lopes, 'Optimization of Image Processing Algorithms for Character Recognition in Cultural Typewritten Documents'.
[83] Hutchinson, 'Natural Language Processing and Machine Learning as Practical Toolsets for Archival Processing'.
[84] Oliver, 'Leveraging KOS to Extend Our Reach with Automated Processes'.

One primary integration issue involves compatibility with existing library and archival infrastructures. The University of Nebraska, for example, experienced significant technical difficulties in harmonizing AI-driven workflows with legacy archival systems, resulting in isolated and non-scalable implementations[85]. Similarly, the British Library, collaborating with the Alan Turing Institute, faced major challenges in embedding tools like Transkribus and BERT-based models due to non-standardized data structures and inadequate API compatibility[86].

Even when technically feasible, integrating AI solutions often disrupts existing institutional workflows. The University of Saskatchewan reported that their adoption of NLP tools initially led to productivity declines due to necessary workflow adaptations and extensive staff retraining[87]. Additionally, the Frick Art Reference Library found integrating automated image classification into curatorial workflows challenging, primarily due to differing interpretative frameworks between human expertise and AI-driven outputs[88].

Furthermore, institutions frequently encountered difficulties integrating AI tools into user-facing services. For instance, Ryerson University Library reported low adoption rates for its AI-powered chatbot due to user skepticism, limited awareness, and discomfort with automated tools[89]. Likewise, the University of Toronto Libraries noted slow uptake of AI-driven recommendation systems, primarily due to trust issues and perceived opacity in algorithmic decision-making processes[90].

Metadata Inconsistency and Interoperability Barriers (25 out of 54 cases, 46%)

Inconsistent metadata standards and interoperability problems represent significant barriers in AI implementation, severely constraining automation and effectiveness in some of the case studies. The German National Library experienced substantial difficulties applying AI-based classification systems across multiple datasets with differing metadata schemas, necessitating extensive manual interventions[91]. The EyCon Project similarly identified significant issues when applying universalized vocabularies and ontologies to culturally specific image archives. This mismatch led to overclassification, misrepresentation, and loss of original contextual meaning[92]. Additionally, the CAMPI initiative (Computer-Aided Metadata Generation for Photo Archives) reported that existing metadata inconsistencies across collections significantly limited their success in integrating automated image analysis tools[93].

Resource Management Challenges (21 out of 54 cases, 39%)

Resource management issues—including shortages in skilled personnel, institutional preparedness, and sustainable infrastructure—present substantial challenges to successful AI implementation in

[85] Hutchinson, ‘Natural Language Processing and Machine Learning as Practical Toolsets for Archival Processing’.
[86] Jaillant, Aske, and Caputo, ‘The National Archives (UK): Case Study’.
[87] Hutchinson, ‘Natural Language Processing and Machine Learning as Practical Toolsets for Archival Processing’.
[88] C Coleman, C Engel, and H Thorsen, ‘Subjectivity and Discoverability: An Exploration with Images’, in *The Rise of AI: Implications and Applications for AI in Academic Libraries. Chicago, IL: Association of College and Research Libraries*, 2022, 83–94.
[89] Ehrenpreis and DeLooper, ‘Implementing a Chatbot on a Library Website’.
[90] Dikow et al., ‘Developing Responsible AI Practices at the Smithsonian Institution’.
[91] Modden, ‘Artificial Intelligence, Machine Learning and Bibliographic Control: DDC Short Numbers: Towards Machine-Based Classifying’.
[92] Aske and Giardinetti, ‘(Mis) Matching Metadata: Improving Accessibility in Digital Visual Archives through the EyCon Project’.
[93] Flannery, ‘Using NLP to Generate MARC Summary Fields for Notre Dame’s Catholic Pamphlets’.

libraries and archives. These barriers significantly affect project viability, operational continuity, and scalability.

A major issue is the shortage of trained personnel equipped with both technical and archival knowledge. The University of Kalyani highlighted significant staffing shortages in data science and AI system management, limiting their ability to scale semi-automated subject indexing systems effectively[94]. Similarly, the National Archives (UK) noted additional operational burdens as staff were required to acquire new technical skills beyond their traditional archival roles, creating significant strain[95].

Sustaining AI-driven projects beyond initial phases proved consistently challenging in the case studies. San Jose State University Library, for instance, discontinued a promising AI chatbot due to escalating resource demands for technical maintenance and upgrades, highlighting broader sustainability concerns[96]. Similarly, North-West University's experimentation with humanoid robotics for library services demonstrated initial technical feasibility but was eventually discontinued due to the high complexity and costs of maintaining the necessary infrastructure[97].

Additionally, inadequate organizational infrastructure significantly restricts effective AI deployment and scaling. The AEOLIAN project found that cultural institutions often lacked critical digital infrastructure such as cloud storage, adequate computing environments, and version control systems required to effectively operate and scale AI-driven solutions[98]. Institutions also frequently experienced unclear governance roles and misalignments between IT departments and digital transformation units, resulting in operational delays and duplicated efforts.

Collectively, these resource management challenges underline the need for strategic institutional investments, targeted training, and improved organizational structures to effectively support sustained AI adoption in library and archival contexts.

Discussion

In this section it is considered how far 1) our analysis of the case studies confirm the AI capability model[99] and 2) whether libraries and archives do have AI capability. First let us consider tangible resources, which is said to consist of data, infrastructure and financial resources. In all three dimensions these did seem central to the case studies. As regards data, libraries and archives possess specialized, historical, and culturally significant datasets (such as texts and historical images). It is the enhanced use of these that is central to their AI implementation efforts. This strongly aligns with the capability model's emphasis on data as foundational for successful AI implementation. Equally issues such as OCR inaccuracies and other data-related problems reported underscore the necessity of structured, high-quality data[100], but do point to some areas of weakness in libraries' AI capability.

As regards the second aspect, the case studies' focus on infrastructure confirms the importance of this resource. However, the extent to which institutions possess strong AI infrastructure varies

---

[94] Ahmed, Mukhopadhyay, and Mukhopadhyay, 'Automated Knowledge Organisation: AI/ML-Based Subject Indexing System for Libraries.'
[95] Jaillant, Aske, and Caputo, 'The National Archives (UK): Case Study'.
[96] Ehrenpreis and DeLooper, 'Implementing a Chatbot on a Library Website'.
[97] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.
[98] Hutchinson, 'Natural Language Processing and Machine Learning as Practical Toolsets for Archival Processing'.
[99] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.
[100] Kim et al., 'Cleaning Dirty Books: Post-OCR Processing for Previously Scanned Texts'.

significantly, with some demonstrating advanced AI capabilities while others rely on external services or remain in developmental phases. The National Library of Sweden, for instance, has developed advanced NLP models like KB-BERT and BERTopic, indicating a strong capability in text processing[101]. However, whether its infrastructure extends to broader AI applications remains unclear. KBR Belgium, on the other hand, effectively utilizes cloud-based AI solutions such as Microsoft Power Automate and AI Builder, demonstrating successful AI adoption but not necessarily indicating a standalone, high-performance AI infrastructure[102]. Meanwhile, Estonia's Kratt project remains in an experimental stage, illustrating AI potential but not yet reflecting a fully developed or integrated AI capability[103]. These examples highlight that while AI is actively employed across institutions, it is difficult to generalize that all of them possess a robust AI infrastructure as defined by the AI capability model.

Mikalef and Gupta[104]'s model emphasizes the need for financial resources. Again, financial resources were often seen as a constraint on what libraries could achieve, confirming the importance of this factor. The case studies indicate significant disparities between institutions in the Global North and Global South, reflecting socio-economic constraints effect AI use in poorer countries[105].

The second type of resource the capability model says is needed for AI capability are human skills, consisting of technical and business skills. The case studies appear to confirm this as an important aspect. AI applications in libraries require significant technical expertise, particularly in initial training, evaluation, and fine-tuning of models. This reliance is highlighted by the fact that while some institutions possess AI-related technical skills (in Machine Learning, NLP, OCR), others experience skill gaps that hinder effective AI implementation, requiring external support or extensive training. Case studies confirm that institutions with well-developed AI capabilities can more effectively implement and optimize AI applications, whereas those lacking in-house expertise struggle with AI integration and long-term sustainability. This variation underscores the ongoing need for technical skill development in the sector to fully align with the model's requirements[106].

The model of AI capability[107] explicitly states that business skills required for AI success include leadership, cross-functional team coordination, and the ability to align AI with organizational goals. Case studies confirm that strong leadership is critical in AI adoption, as institutions that lacked leadership support struggled to implement AI effectively. Additionally, the model highlights that more than 40% of organizations face cultural resistance to AI, which hinders adoption and business value[108]. This underscores the necessity of business skills such as change management and strategic decision-making, particularly in ensuring staff buy-in and aligning AI projects with institutional goals[109]. Furthermore, cross-departmental collaboration is emphasized as essential for bridging gaps

---

[101] Martin Malmsten, Love Börjeson, and Chris Haffenden, 'Playing with Words at the National Library of Sweden--Making a Swedish BERT', *arXiv Preprint arXiv:2007.01658*, 2020.
[102] Hannes Lowagie, 'One Automatic Cataloging Flow: Tests and First Results', *International Federation of Library Associations and Institutions (IFLA)*, 2023, https://repository.ifla.org/server/api/core/bitstreams/5f1236c0-23e2-4ba9-b195-4efd19c01bff/content.
[103] Asula et al., 'Kratt: Developing an Automatic Subject Indexing Tool for the National Library of Estonia'.
[104] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.
[105] Ahmed, Mukhopadhyay, and Mukhopadhyay, 'Automated Knowledge Organisation: AI/ML-Based Subject Indexing System for Libraries.'; Asula et al., 'Kratt: Developing an Automatic Subject Indexing Tool for the National Library of Estonia'.
[106] Malmsten, Börjeson, and Haffenden, 'Playing with Words at the National Library of Sweden--Making a Swedish BERT'; Asula et al., 'Kratt: Developing an Automatic Subject Indexing Tool for the National Library of Estonia'.
[107] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.
[108] Mikalef and Gupta.
[109] Jaillant, Aske, and Caputo, 'The National Archives (UK): Case Study'.

between technical teams and non-technical stakeholders, as seen in AI initiatives across several institutions[110]. Case studies such as Oliver[111] support the AI capability model's assertion that institutions must invest in business skills to navigate challenges like resistance to change, interdepartmental misalignment, and a lack of strategic oversight. Cross-departmental collaboration and communication are seen as necessary to bridge the gap between technical teams and non-technical stakeholders.

The third type of resource in the model are intangible resources. The role of interdisciplinary collaboration, involving teams of archivists, librarians, digital humanists, and computer scientists, aligns closely with the AI capability model's emphasis on cross-functional coordination. Mikalef & Gupta[112] identify interdepartmental coordination as a key intangible resource, enabling effective AI adoption through communication and shared goals across teams. This is reflected in case studies such as Dikow et al.[113], where collaboration between curators, data scientists, and biologists supported machine learning applications in natural history collections. Similarly, Cornut et al.[114] highlight interdisciplinary cooperation in the development of AI annotation tools involving computer scientists and archival professionals.

Mikalef and Gupta[115] stress risk proclivity as a key intangible resource. Libraries and archives generally adopt a cautious, incremental approach to AI implementation rather than taking bold, high-risk initiatives. Case studies suggest that this approach is influenced by financial constraints, ethical considerations, and regulatory oversight, which prioritize risk mitigation over rapid AI adoption[116]. While the model suggests that organizations embracing risk-taking gain a competitive edge, cultural institutions appear to prioritize sustainability and careful integration over aggressive AI expansion[117]. This can be seen as an area of limited capability.

Thus in many aspects the close analysis of the case studies supports many aspects of the model of capability put forward by Mikalef and Gupta[118]. However, there are several factors that appeared to be very important in our case study collection not mentioned in the model that perhaps need to be added. They often emerged in discussing areas where libraries were lacking. They can be summarised in three areas: ethical concerns; AI literacy for all staff; and Interoperability. While technical expertise plays a role in bias mitigation, broader ethical concerns are central to how AI can be applied in the library and archive context. This gap aligns to Freeman et al.[119]'s point that the RBV neglects values and ethics. Solving these issues often seem to involve governance frameworks and institutional

---

[110] Dikow et al., 'Developing Responsible AI Practices at the Smithsonian Institution'; Cornut, Raemy, and Spiess, 'Annotations as Knowledge Practices in Image Archives: Application of Linked Open Usable Data and Machine Learning'.
[111] Oliver, 'Leveraging KOS to Extend Our Reach with Automated Processes'.
[112] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.
[113] Dikow et al., 'Developing Responsible AI Practices at the Smithsonian Institution'.
[114] Cornut, Raemy, and Spiess, 'Annotations as Knowledge Practices in Image Archives: Application of Linked Open Usable Data and Machine Learning'.
[115] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.
[116] Jaillant, Aske, and Caputo, 'The National Archives (UK): Case Study'.
[117] Jaillant, Aske, and Caputo.
[118] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.
[119] Freeman, Dmytriyev, and Phillips, 'Stakeholder Theory and the Resource-Based View of the Firm'.

policies[120], areas underrepresented in the AI Capability Model. But ethical skills could also be seen as an aspect of human skills not in the model.

Effective deployment of AI consistently required substantial training and digital literacy programs for staff, aligning with Mikalef and Gupta[121]'s model, which highlights the need for technical expertise. However, AI literacy is not just necessary for technical personnel—it is crucial across the entire organization. Case studies show that non-technical staff, such as librarians and managers, also require AI awareness and training to ensure informed decision-making, ethical implementation, and effective collaboration between departments[122]. So again one could argue that the wide diffusion of AI skills across the organization should be stressed as an essential resource for AI capability.

Thirdly, integration and interoperability posed technical and structural challenges. The Kratt tool in Estonia faced difficulties integrating into cataloguing workflows[123]. At the British Library, integration of tools like Transkribus and BERT based models was hindered by legacy infrastructure and non-standardized data structures[124]. Similarly, metadata inconsistencies across collections complicated AI deployment at the German National Library[125]. As these examples illustrate the integration of new AI based services with legacy systems seems a key dimension of AI capability. This is less a weakness of the capability model and more a reflection that AI capability is just part of the organisation's overall operations.

Conclusion

This paper has sought to evaluate Mikalef and Gupta[126]'s AI capability model in the library and archive context. Through the content analysis of published case studies of library and archive applications of AI, it asks whether the model appears to fully represent the resources needed to implement AI. It further asks if by this criteria libraries and archives do have AI capability and, if not, where the key gaps are. A number of important resources not mentioned in the model are uncovered, particularly the need for ethical skills, the need for AI awareness across the whole organisation and the challenges of integrating AI with legacy systems. However, in general it seems that the case studies confirm many aspects of the model.

The data of the case studies suggests that there are many gaps in library and archive AI capability, but especially in areas such as lack of infrastructure and technical resources, data issues arising from metadata inconsistencies, and lack of financial resources. Many projects struggle with metadata inconsistencies, lack of standardization, and data sparsity, which hinder AI training and classification accuracy. Financial constraints also shaped outcomes, particularly in the Global South, where institutions faced more difficulty developing infrastructure or hiring skilled staff. In terms of skills, access to the necessary technical and managerial skills varied. More broadly, institutions cited difficulty retaining skilled staff and managing complex tool configurations. In terms of intangible resources: resistance to change, lack of leadership, and limited coordination were common. For

[120] Benjamin Charles Germain Lee et al., 'The Newspaper Navigator Dataset: Extracting and Analyzing Visual Content from 16 Million Historic Newspaper Pages in Chronicling America', *arXiv Preprint arXiv:2005.01583*, 2020; Dikow et al., 'Developing Responsible AI Practices at the Smithsonian Institution'.
[121] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.
[122] Helen Sau Ching Cheung et al., '21 Developing Digital Literacy Using Mini-AI Games', in *New Horizons in Artificial Intelligence in Libraries* (Walter de Gruyter GmbH & Co KG, 2024).
[123] Asula et al., 'Kratt: Developing an Automatic Subject Indexing Tool for the National Library of Estonia'.
[124] Lincoln et al., 'CAMPI: Computer-Aided Metadata Generation for Photo Archives Initiative', 2020.
[125] Lowagie, 'One Automatic Cataloging Flow: Tests and First Results'.
[126] Mikalef and Gupta, 'Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance'.

readers of the journal it is hoped that this exploration of AI capability sheds direct light on the capability to apply AI to cataloguing and classification tasks, especially in their broader context of an ecosystem of tools and organisational structures. It is unlikely that libraries can have the capacity to introduce AI for tasks such as metadata creation, without consideration of the level of wider capability.

An important reflection on this analysis is that the case studies should not be taken to represent typical activity in libraries and archives, in two ways. Firstly, it is probable that the libraries reporting their activities in published reports are unrepresentative. They are likely to be the more privileged, high resourced institutions, such as national libraries and wealthier academic libraries. Their activity is more advanced than is typical. Secondly, published papers are much more likely to report the most innovative activities because they are new to the literature. More familiar uses that are becoming standard practice will be mentioned less. Taking this reflection into account would suggest that the position in libraries and archives in general may be less advanced than portrayed in our paper. If anything the case studies over represent library and archive capability. What the paper has achieved is providing supporting evidence that the AI capability model proposed by Mikalef and Gupta[127]'s is useful in evaluating libraries' ability to implement AI, at least if extended with the three resources mentioned at the end of the discussion.

Of course, how AI is implemented into library and archive services is not the only narrative around AI. A large part of the response of libraries to AI, particularly post ChatGPT, revolves around raising AI literacy, rather than applying AI to library services. Nevertheless, the focus on how AI is implemented into library services does represent an important part of the story about AI and libraries and archives. As more data emerges about libraries' and archives' use of AI, future research will be able to offer more comprehensive evaluations of trends, tools, and capability development.

[127] Mikalef and Gupta.

Appendix 1

Ahmed, Mustak, Mondrita Mukhopadhyay, and Parthasarathi Mukhopadhyay. 'Automated Knowledge Organisation: AI/ML-Based Subject Indexing System for Libraries.' *DESIDOC Journal of Library & Information Technology* 43, no. 1 (2023).

Akça, Sümeyye. '15 Topic Modelling in the Ottoman Kadi Registers'. In *New Horizons in Artificial Intelligence in Libraries*. Walter de Gruyter GmbH & Co KG, 2024.

Aske, Katherine, and Marina Giardinetti. '(Mis) Matching Metadata: Improving Accessibility in Digital Visual Archives through the EyCon Project'. *ACM Journal on Computing and Cultural Heritage* 16, no. 4 (2023): 1–20.

Asula, Marit, Jane Makke, Linda Freienthal, Hele-Andra Kuulmets, and Raul Sirel. 'Kratt: Developing an Automatic Subject Indexing Tool for the National Library of Estonia'. *Cataloging & Classification Quarterly* 59, no. 8 (2021): 775–93.

Averkamp, Shawn, Kerri Willette, Amy Rudersdorf, and Meghan Ferriter. 'Humans-in-the-Loop Recommendations Report', 2021.

Balnaves, Edmund. '23 Artificial Intelligence in Libraries on $5 per Day: Image Matching with Koha'. In *New Horizons in Artificial Intelligence in Libraries*, edited by Edmund Balnaves, Leda Buldrini, Andrew Cox, and Raymond Uzwyshyn, 2024. https://doi.org/10.1515/9783111336435.

Brygfjeld, Svein Arne, Freddy Wetjen, and André Walsøe. 'Machine Learning for Production of Dewey Decimal'. Kuala Lumpur, 2018. https://library.ifla.org/id/eprint/2216/1/115-brygfjeld-en.pdf.

Cheung, Helen Sau Ching, Alex Hok Lam Chan, Kenny Ka Lam Kwan, and Yoko Hirose. '21 Developing Digital Literacy Using Mini-AI Games'. In *New Horizons in Artificial Intelligence in Libraries*. Walter de Gruyter GmbH & Co KG, 2024.

Chou, Charlene, and Tony Chu. 'An Analysis of BERT (NLP) for Assisted Subject Indexing for Project Gutenberg'. *Cataloging & Classification Quarterly* 60, no. 8 (2022): 807–35.

Coleman, C, C Engel, and H Thorsen. 'Subjectivity and Discoverability: An Exploration with Images'. In *The Rise of AI: Implications and Applications for AI in Academic Libraries. Chicago, IL: Association of College and Research Libraries*, 83–94, 2022.

Coleman, Catherine Nicole. 'Computer Vision and Cultural Heritage Case Study 2'. Stanford: AEOLIAN, 2021. https://www.aeolian-network.net/wp-content/uploads/2022/04/Aeolian-Case-Study-2-April-2022.pdf.

Cornut, Murielle, Julien Antoine Raemy, and Florian Spiess. 'Annotations as Knowledge Practices in Image Archives: Application of Linked Open Usable Data and Machine Learning'. *ACM Journal on Computing and Cultural Heritage* 16, no. 4 (2023): 1–19.

Dias, Mariana, and Carla Teixeira Lopes. 'Optimization of Image Processing Algorithms for Character Recognition in Cultural Typewritten Documents'. *ACM Journal on Computing and Cultural Heritage* 16, no. 4 (2023): 1–25.

Dikow, Rebecca, Corey DiPietro, Michael Trizna, Hanna BredenbeckCorp, Madeline Bursell, Jenna Ekwealor, Richard Hodel, Nilda Lopez, William Mattingly, and Jeremy Munro. 'Developing Responsible AI Practices at the Smithsonian Institution'. *ARPHA Preprints* 4 (2023): e113335.

Drobac, Senka, Pekka Kauppinen, and Krister Linden. 'Improving OCR of Historical Newspapers and Journals Published in Finland'. Brussels: DataTech, 2019.

Duong, Quan, Mika Hämäläinen, and Simon Hengchen. 'An Unsupervised Method for OCR Post-Correction and Spelling Normalisation for Finnish'. *arXiv Preprint arXiv:2011.03502*, 2020.

Ehrenpreis, Michelle, and John DeLooper. 'Implementing a Chatbot on a Library Website'. *Journal of Web Librarianship* 16, no. 2 (2022): 120–42.

Enis, Enis, Matt. 'AI And The Public: As Artificial Intelligence Becomes More Embedded In Work, Creative Pursuits, And The Generation Of Online Misinformation, Public Libraries Have A Major New Role To Play In Digital Literacy'. *Proquest*, 2024.

Enis, Matt. 'Library of Congress Trains Machine Learning with Crowdsourcing'. *Library Journal*, 2022. https://www.libraryjournal.com/story/library-of-congress-trains-machine-learning-tool-with-crowdsourcing.

Flannery, Jeremiah. 'Using NLP to Generate MARC Summary Fields for Notre Dame's Catholic Pamphlets'. *International Journal of Librarianship* 5, no. 1 (2020): 20–35.

Golub, Koraljka. 'Automated Subject Indexing: An Overview'. *Cataloging & Classification Quarterly* 59, no. 8 (2021): 702–19.

Gupta, Varun. 'From Hype to Strategy: Navigating the Reality of Experimental Strategic Adoption of AI Technologies in Libraries'. *Reference Services Review* 53, no. 1 (2024): 1–14.

Haffenden, Chris, Elena Fano, Martin Malmsten, and Love Börjeson. 'Making and Using AI in the Library: Creating a BERT Model at the National Library of Sweden'. *College & Research Libraries* 84, no. 1 (2023).

Hutchinson, Tim. 'Natural Language Processing and Machine Learning as Practical Toolsets for Archival Processing'. *Records Management Journal* 30, no. 2 (2020): 155–74.

Jaillant, Lise, Katherine Aske, and Annalina Caputo. 'The National Archives (UK): Case Study'. AEOLIAN, 2021.

Jiang, Ming, Yuerong Hu, Glen Worthey, Ryan C Dubnicek, Ted Underwood, and J Stephen Downie. 'Impact of OCR Quality on BERT Embeddings in the Domain Classification of Book Excerpts'. *Proceedings Http://Ceur-Ws. Org ISSN* 1613 (2021): 0073.

Kasprzik, Anna. '14 Transferring Applied Machine Learning Research into Subject Indexing Practice'. In *New Horizons in Artificial Intelligence in Libraries*, Vol. 185. Walter de Gruyter GmbH & Co KG, 2024.

Kettunen, Kimmo, Mika Koistinen, and Jukka Kervinen. 'Ground Truth OCR Sample Data of Finnish Historical Newspapers and Journals in Data Improvement Validation of a Re-OCRing Process'. *LIBER Quarterly: The Journal of the Association of European Research Libraries* 30, no. 1 (2020): 1–20.

Kim, Allen, Charuta Pethe, Naoya Inoue, and Steve Skiena. 'Cleaning Dirty Books: Post-OCR Processing for Previously Scanned Texts'. *arXiv Preprint arXiv:2110.11934*, 2021.

Kluge, Lisa, and Maximilian Kahler. 'Few-Shot Prompting for Subject Indexing of German Medical Book Titles'. In *Proceedings of the 20th Conference on Natural Language Processing (KONVENS 2024)*, 141–48. Vienna: Association for Computational Linguistics, 2024.

Kummervold, Per E, Javier De la Rosa, Freddy Wetjen, and Svein Arne Brygfjeld. 'Operationalizing a National Digital Library: The Case for a Norwegian Transformer Model'. In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, 20–29. Reykjavik: Linkoping University Electronic Press, 2021. https://aclanthology.org/2021.nodalida-main.3/.

Lee, Benjamin. ‘Compounded Mediation: A Data Archaeology of the Newspaper Navigator Dataset’. *Digital Humanities Quarterly* 15, no. 4 (2021): 3055–62.

Lee, Benjamin Charles Germain, Jaime Mears, Eileen Jakeway, Meghan Ferriter, Chris Adams, Nathan Yarasavage, Deborah Thomas, Kate Zwaard, and Daniel S Weld. ‘The Newspaper Navigator Dataset: Extracting and Analyzing Visual Content from 16 Million Historic Newspaper Pages in Chronicling America’. *arXiv Preprint arXiv:2005.01583*, 2020.

Lincoln, Matthew, Julia Corrin, Emily Davis, and Scott B Weingart. ‘CAMPI: Computer-Aided Metadata Generation for Photo Archives Initiative’, 2020.

Lorang, Elizabeth, Leen-Kiat Soh, Yi Liu, and Chulwoo Pack. ‘Digital Libraries, Intelligent Data Analytics, and Augmented Description: A Demonstration Project’, 2020.

Lowagie, Hannes. ‘One Automatic Cataloging Flow: Tests and First Results’. *International Federation of Library Associations and Institutions (IFLA)*, 2023. https://repository.ifla.org/server/api/core/bitstreams/5f1236c0-23e2-4ba9-b195-4efd19c01bff/content.

Malmsten, Martin, Love Börjeson, and Chris Haffenden. ‘Playing with Words at the National Library of Sweden--Making a Swedish BERT’. *arXiv Preprint arXiv:2007.01658*, 2020.

Malmsten, Martin, Chris Haffenden, and Love Börjeson. ‘Hearing Voices at the National Library--a Speech Corpus and Acoustic Model for the Swedish Language’. *arXiv Preprint arXiv:2205.03026*, 2022.

Malmsten, Martin, Viktoria Lundborg, Elena Fano, Chris Haffenden, Fredrik Klingwall, Robin Kurtz, Niklas Lindström, Faton Rekathati, and Love Börjeson. ‘Without Heading? Automatic Creation of a Linked Subject System’, 2024.

Mendoza, Sonia, Luis Martín Sánchez-Adame, José Fidel Urquiza-Yllescas, Beatriz A González-Beltrán, and Dominique Decouchant. ‘A Model to Develop Chatbots for Assisting the Teaching and Learning Process’. *Sensors* 22, no. 15 (2022): 5532.

Modden, Elisabeth. ‘Artificial Intelligence, Machine Learning and Bibliographic Control: DDC Short Numbers: Towards Machine-Based Classifying’. *JLIS: Italian Journal of Library, Archives and Information Science= Rivista Italiana Di Biblioteconomia, Archivistica e Scienza Dell’informazione: 13, 1, 2022*, 2022, 256–64.

Nie, Bing, Ting Wang, Brady Daniel Lund, and Fengping Chen. ‘How Does Ai Make Libraries Smart?: A Case Study of Hangzhou Public Library’. In *Technological Advancements in Library Service Innovation*, 43–58. IGI Global Scientific Publishing, 2022.

Oliver, Chris. ‘Leveraging KOS to Extend Our Reach with Automated Processes’. *Cataloging & Classification Quarterly* 59, no. 8 (2021): 868–74.

Prokop, Ellen, XY Han, Vardan Papyan, David L Donoho, and C Richard Johnson Jr. ‘AI and the Digitized Photoarchive: Promoting Access and Discoverability’. *Art Documentation: Journal of the Art Libraries Society of North America* 40, no. 1 (2021): 1–20.

Ramakrishnan, Ritu, Priyanka Thangamuthu, Austin Nguyen, and Jinzhu Gao. ‘Revolutionizing Campus Communication: NLP-Powered University Chatbots.’ *International Journal of Advanced Computer Science & Applications* 15, no. 6 (2024).

Reinsfelder, Thomas L, and Katie O’Hara-Krebs. ‘Implementing a Rules-Based Chatbot for Reference Service at a Large University Library’. *Journal of Web Librarianship* 17, no. 4 (2023): 95–109.

Rodriguez, Sharesly, and Christina Mune. ‘Uncoding Library Chatbots: Deploying a New Virtual Reference Tool at the San Jose State University Library’. *Reference Services Review* 50, no. 3/4 (2022): 392–405.

Stevenson, Adrian. ‘Employing Machine Learning and Artificial Intelligence in Cultural Institutions’. The Archives Hub, 2021.

Toane, Carey, Lise Doucette, Paulina Rousseau, Michael Serafin, Michelle Spence, and Christina Kim. ‘The 99 AI Challenge: Empowering a University Community through an Open Learning Pilot’. In *The Rise of AI: Implications and Applications for AI in Academic Libraries*, edited by Sandy Hervieux and Amanda Wheatley. Association of College and Research Libraries, 2022. http://hdl.handle.net/1807/111244.

Trehub, Aaron, and Ali Krzton. ‘Using IBM Watson for Discovery and Research Support: A Library-Industry Partnership at Auburn University’. In *The Rise of AI: Implications and Applications of Artificial Intelligence in Academic Libraries*, edited by Sandy Hervieux and Amanda Wheatley, 149–61. ACRL, 2022.

Tshabalala, Neli. ‘10 Impact of Artificial Intelligence on Library Services: Reflections on a Practical Project’. In *New Horizons in Artificial Intelligence in Libraries*, edited by Edmund Balnaves, Leda Buldrini, Andrew Cox, and Raymond Uzwyshyn. Walter de Gruyter GmbH & Co KG, 2024. doi.org/10.1515/9783111336435-011.

Twomey, Beth, Annie Johnson, and Colleen Estes. ‘It Takes a Village: A Distributed Training Model for AI-Based Chatbots’ 43, no. 3 (2024). https://doi.org/10.5860/ital.v43i3.17243.

Wang, F, A Tucker, and J Seo. ‘Incubating AI: The Collaboratory at Ryerson University Library’. In *The Rise of AI: Implications and Applications for AI in Academic Libraries. Chicago, IL*, 47–60, 2022.

Zaragoza, Thomas, Yann Nicolas, and Aline Le Provost. ‘From Text to Data inside Bibliographic Records. Entity Recognition and Entity Linking of Contributors and Their Roles from Statements of Responsibility’. In *New Horizons in Artificial Intelligence in Libraries*. Walter de Gruyter GmbH & Co KG, 2024.