

# Ethical AI in Libraries: Governance, Transparency, and Bias Mitigation in Discovery and Metadata Systems

Chennumalla Srilakshmi

*IGNOU M.Li.Sc.*

**Abstract-** The growing incorporation of artificial intelligence (AI) in library discovery services, recommendation services, and metadata mechanisms has fostered the improvement of information search and customized user services. Nevertheless, the integration of AI is already accompanied with some essential ethical issues, especially concerning governance, transparency, and bias reduction measures in AI-based discovery and metadata systems, this paper will synthesize existing scholarly literature, policy reports and professional guidelines. This paper presents a narrative review of the literature based on a qualitative secondary data approach to examine how much the current AI implementations in libraries follow ethical principles of fairness, accountability, explain ability, and privacy as defined in such frameworks as the Ethically Aligned Design by IEEE, the AI Act in the EU, and the AI Principles by the OECD. The review finds the various enduring challenges in the opaque vendor-controlled algorithms, biased training data, and weak institutional governance structures. By systematizing ethical issues and practices, the present study can enhance the understanding of ethical concerns in the adoption of AI in libraries and provide policy and practice-oriented recommendations to promote just, transparent, and trustful AI-enabled library services.

**Keywords:** Ethical Artificial Intelligence, Library Discovery Systems, Algorithmic Bias, AI Governance, Metadata Ethics

## I. INTRODUCTION

### Background

The use of artificial intelligence (AI) in libraries has developed at an exceptionally high pace, transforming the way tools of discovery, recommendation engines, and metadata enrichment are performed. Libraries are starting to use AI technologies to improve search provisions, deal with big and intricate collections and provide personalized services to clients and increase

the efficiency and user experience of the operation (Balnaves, 2024; Halburagi and Mukarambi, 2023). These advances are part of overall technological trends in information management dysfunctional systems are used to facilitate automated cataloguing, predictive analytics to build collection development, and dynamic recommendation services based on the user requirements of customers (Manjunatha, 2023; Oyighan et al., 2024).

### Problem Statement

However, the increasing application of AI in libraries setting provokes serious ethical issues, notwithstanding all these advantages. Issues such as poor governance frameworks, poor transparency in algorithmic decision-making, and discovery and metadata systems bias are the primary challenges. Algorithms exhibit biases that are usually acquired via biased training data or historically entrenched classification processes and thus lead to unequal access to resources and inhibit the realization of equitable access to information. The end results of such outcomes go against the basic tenets of libraries, such as inclusivity, intellectual freedom, and trust in the users and thus require systematic ethical governance.

### Purpose and Scope

The paper is a synthesis of literature available on the topic of ethical AI in libraries, particularly governance systems, transparency systems, and the methodology of reducing algorithmic bias. Ethical risks and impacts are the most evident in the area of AI-enabled discovery systems, including next-generation library catalogs and recommender systems, and automated metadata generation and management.

### Research Questions

The research questions guiding the study entail:

1. Which ethical principles of AI application are aided by their governing structures in the library setting?
2. How clear and explicable are AI systems in libraries?
3. Is there any evidence of algorithmic bias in library discovery systems and metadata systems?
4. What does the literature recommend as strategies and best practices to deal with these challenges?

#### Significance of the Study

The article can be used to advance the current debates on ethical AI in library and information science by synthesizing the results of available research and policy frameworks. The results provide both practical and theoretical suggestions to librarians, system designers, and policymakers in adopting AI responsibly so that the library services can be fair, equitable, and accountable.

## II. THEORETICAL AND CONCEPTUAL FRAMEWORK

### Defining Ethical Artificial Intelligence

Ethical artificial intelligence (AI) is a concept that describes the process of designing, developing, and implementing AI systems in a manner that optimizes the harm reduction and social benefit maximization (Balnaves, 2024; Korošec, 2024; Steinerová and Ondrišova, 2024). In the most significant ethical theories, fairness, accountability, transparency, explainability, and privacy are always the primary principles. Fairness aims at stopping the results of discrimination and provides fair treatment; accountability means that it is necessary to hold someone accountable to make decisions based on AI; transparency and explainability allow people to comprehend and examine the algorithm processes; and privacy protects personal and sensitive user data (Balnaves, 2024; Korošec, 2024; Sontan and Samuel, 2024). All these principles manage the risks of amplification of bias, non-transparency, and loss of user confidence, creating the basis of trustworthy AI

in institutional settings, including libraries (Samek and Shiri, 2022; Steinerova and Ondrišova, 2024).

### Conceptualization of Ethical AI in Libraries

The paper defines ethical AI in libraries as a nexus of three mutually reinforcing aspects: governance, transparency, and bias mitigation approaches, which in combination define the ethical behavior of the AI-based discovery and metadata systems. Governance will offer the structural and policy-based regulation needed to direct AI deployment; transparency will allow interpretation of the algorithm processes and accountability, bias reduction will be able to provide fair representation and access to information. These dimensions do not exist in isolation but instead are reinforcing since good governance will support transparency demands, transparency supporting bias detection and bias reduction supporting ethical accountability of library systems.

### Relevant Ethical AI Frameworks

#### General Technology Frameworks

The developed AI ethics presented by larger technology industries can be used as the base of recommendations that can be employed in libraries. The Ethically Aligned Design framework created by the IEEE focuses on human rights, well-being, data agency, transparency, accountability, and understanding of misuse (Khan et al., 2022). Equally, the AI Act by European Union recommends a risk-based system of regulation that ensures fairness, transparency, accountability, and privacy, but bans risky and manipulative AI use (Butt, 2024). The OECD AI Principles also embrace the inclusive growth, human considerations, robustness, transparency, and accountability and have become common in national policy settings (Abercrombie et al., 2024; Nasr-Azadani and Chatelain, 2024).

#### Adaptations to Library and Information Science

In the context of library and information science, these broad structures are modified in line with the main professional values, including intellectual freedom, inclusivity, equity, and trust with the user. Ethical principles presented in library scholarship focus on the human control, the regulation of privacy, the transparency, and equity in the AI-powered systems

(Balnaves, 2024; Korošec, 2024; New Horizons in Artificial Intelligence in Libraries, 2024). The principles guiding research libraries also promote the formal system of AI governance with the focus on transparency and mitigation of bias in the discovery and access tools (Libraries, 2024). Bibliometric research in the information ethics points at an increasing trend of the transition to AI ethics in LIS, emphasizing the significance of data justice and algorithm literacy in the work of library professionals (Samek and Shiri, 2022; Steinerova and Ondrišova, 2024).

### Key Concepts in the Library Context

#### AI-Driven Discovery Systems

Semantic search and relevance ranking Improved information retrieval The AI-based discovery systems include next-generation library catalogs and single search platforms. These systems, however, have the risk of replicating the biases bestowed on the traditional classification schemes and automated processes of text analysis (Berendt et al., 2023; Wilson, 2020). To achieve fairness in discovery, there is need to systematically audit representational and ranking biases to facilitate a variety of fair access to library resources (Berendt et al., 2023; Igbinovia and Mensah, 2025).

#### Metadata Creation, Curation, and Enrichment

Metadata generation and enrichment, especially of large digital collections is now increasingly automated using AI. The transparency of machine-generated records, ongoing bias detection, and the human validation of records are thus some of the ethical metadata practices that ensure accuracy, inclusivity and contextual integrity (Korošec, 2024; Oyighan et al., 2024).

#### Recommender Systems and Personalization

Recommender systems make library content more engaging and discoverable by personalizing access to library resources, and improve access to similar content due to user behavior and similarity in content. Meanwhile, these systems have ethical issues concerning filter bubbles, ranking discrimination, and privacy threats (Ekstrand et al., 2019, 2022). Within

the library setting, the metrics of fairness should consider a number of stakeholders, the personalization should not come at the expense of non-discriminatory access and user privacy (Ekstrand et al., 2019, 2022).

### III. AI ADOPTION IN LIBRARIES: CURRENT STATE

#### Historical Development

The information systems used by libraries have significantly changed over the years and started with manual card catalogs in the pre-digital age and moved on to online public access catalogs (OPACs) that became operable in the 1980s, allowing them to search by key words (Manjunatha, 2023). The discovery layers of Primo and Summon, which combined federated search of a number of databases and digital collections, appeared in the early 2000s. Most of these systems have been improved over the last few years, with artificial intelligence, such as machine learning (ML) and natural language processing (NLP), added to aid semantic search, predictive analytics, and personalized discovery services (Igbinovia and Mensah, 2025; Oyighan et al., 2024). Although these innovations help to overcome the limitations related to information overload and the intricacy of digital collections, they also pose some threats to the ethical aspects, especially the danger of bias increase due to the use of historically based classification processes and the old data (Berendt et al., 2023).

#### Types of AI Technologies Used

##### Machine Learning for Classification, De-duplication, and Subject Analysis

The use of machine learning methods in bibliographic classification, de-duplication and subject analysis of library catalogs is commonplace. Trained algorithms of the historical metadata enable large-scale classification with systems like the Dewey Decimal Classification and Library of Congress Classification, which are more efficient and consistent as well as reduction of manual labor (Igbinovia and Mensah 2025; Oyighan et al., 2024). Nevertheless, the use of historical training data can also be a continuation of prior biases and this highlights the necessity of ethical considerations when performing automated classification tasks.

**Natural Language Processing for Metadata Extraction**  
 Natural language processing has found greater application in extracting entities, topics and contextual features of unstructured textual materials, such as e-books, journal articles and digitized archival materials. Metadata enrichment can be done at scale with the help of techniques like named entity recognition and topic modeling, and better resources can be discovered (Berendt et al., 2023). Simultaneously, the models of NLP that have been trained on skewed linguistic and culturally biased corpora can recreate bias in metadata representation, especially in relation to non-dominant languages and marginalized areas of knowledge (Wilson, 2020).

#### Recommendation Algorithms

Libraries use collaborative filtering, content-based filtering as well as hybrid system recommendation in their libraries to recommend resources using user behavior, content similarity and usage patterns. Intuitively aligned with commercial recommender systems, library-based recommenders are structured along the lines of facilitating intellectual enquiry and accidental finding, as opposed to consumption (Ekstrand et al., 2019, 2022). Hybrid models propose a solution to such dilemmas as the cold-start problem but create ethical concerns in the prioritization of bias, personalization implications and privacy of the users.

#### Institutional Adoption Patterns

The rates of AI implementation in various types of library institutions are different. Academic libraries are regarded as one of the first to adopt using AI-enhanced discovery systems and analytics tools, those applied to Ex Libris Alma, in support of research-intensive research and digital scholarship efforts (Libraries, 2024; Top Trends in Academic Libraries, 2022). In public libraries, risky with access and equity The focus puts more emphasis on user-facing apps, such as chatbots and personalized discovery interfaces, and balances carefully equity, accessibility, and budgets (Manjunatha, 2023). The national libraries, in contrast, focus on extensive digitization, preservation, and cultural heritage projects and use AI to make multilingual metadata and analyse historic collections automatically (Korošec, 2024).

#### IV. METHODOLOGY

The research design embraced in this study is qualitative and descriptive-analytical through the narrative review that uses secondary data to address the research problem of ethical challenges in artificial intelligence when applied to library discovery systems and metadata systems. Peer-reviewed journals, scholarly books, conference proceedings, and policy documents found in such databases as Scopus, Web of Science, Google Scholar, IEEE Xplore, and leading academic publishers have been searched to find the relevant literature. This was done by a structured keyword search conducted on the keywords related to the topics of ethical AI and library discovery system, metadata, governance, transparency, explainability, and algorithmic bias. The sources were selected as they were published in the year between 2018 and 2025 depending on their relevance to AI applications and the ethical concerns in library and information science.

The thematic and content analysis were applied to the chosen literature based on the conceptual framework of the study that concentrated on AI governance, transparency and explainability and mitigating algorithmic bias. Systematic coding and comparison across sources were the major ways of identifying key themes, patterns, and ethical challenges. The triangulation of scholarly research, professional guideline and policy frameworks, including the EU AI Act and IEEE Ethically Aligned Design, were used to improve the analytical rigor. Although the article is constrained by the use of secondary data and lack of access to proprietary AI systems, it offers a consistent and theoretically informed overview of the ethical AI practices that apply to the modern library setting.

#### V. GOVERNANCE OF AI IN LIBRARIES

##### Policy and Regulatory Landscape

Libraries are also in need of effective frameworks of governance to inform ethical use of AI in accordance to professional values like intellectual freedom, privacy, and equity. The internal library policies are at the center of fostering accountability, establishing what is acceptable use of AI, and adopting ethical control within the deployment and assessment of the system (Libraries, 2024). Such institutional policies

are sometimes guided by more general ethical principles, such as European Commission guidelines that stress on human control, equity, and ethical management of data (Balnaves, 2024; Korošec, 2024; New Horizons in Artificial Intelligence in Libraries, 2024).

At the macro level, AI policies at the national and global level have a big impact on the practices in libraries. The AI act of the European Union takes a risk-based stance and demands the implementation of transparency, accountability, and privacy safeguards and bans manipulative and high-risk AI applications deemed to be relevant in discovery and access systems (Butt, 2024). On the same note, human values, strength, and responsibility are encouraged by the OECD AI Principles, which influence national variations in policy in various jurisdictions (Abercrombie et al., 2024; Nasr-Azadani and Chatelain, 2024). Ethical duties in library data setting continue to be strengthened by the demands of AI policies that can be used to the advantage of humankind put forth by UNESCO (Samek and Shiri, 2022).

#### Governance Models and Responsibilities

The most common governance models of AI in libraries are centralized, decentralized, or hybrid. The former is centralized solutions that put the control in the hands of library leaders to maintain uniform standards and adherence to policies whereas the latter model is decentralized that gives more flexibility and innovativeness to IT units and functional departments. Hybrid models that are increasingly becoming popular in research libraries combine strategic coordination and operational responsiveness (Libraries, 2024).

Effective governance must have clear delineation of roles. The strategic direction and ethical alignment are left to the leadership of the library, technical implementation, monitoring, and evaluation to IT units, and domain expertise and user advocacy should be left to the librarians. Although AI-enabled systems are supplied by vendors, it is the libraries that bear the ethical responsibility of implementing these systems and should proactively address vendor concerns on transparency, auditing, and accountability as a way to

avoid reinforcing bias (Korošec, 2024; New Horizons in Artificial Intelligence in Libraries, 2024).

#### Standards and Professional Guidelines

AI applications are becoming more and more interoperable with the library metadata formats, including MARC, BIBFRAME and Dublin Core, in terms of automated metadata generation and enrichment. Ethical guidelines to AI governance in libraries are also given by professional bodies. The American Library Association (ALA) and the International Federation of Library Associations and Institutions (IFLA) focus on the ideas of equality, inclusiveness, and intellectual freedom, whereas the Ethically Aligned Design framework by IEEE provides a wider human-rights-based view on openness and responsibility (Khan et al., 2022). Libraries modify those instructions to ensure that professionals become more algorithmically literate and informed in making their decisions (Samek and Shiri, 2022; Steinerova and Ondrishova, 2024).

#### Gaps and Ongoing Challenges

Although becoming more conscious of the issue, most libraries have not enacted extensive and institutionalized AI governance systems, which puts them at risk of unregulated bias and unequal ethical standards. Also, dependence on proprietary systems with vendors having control over them restricts institutional observation of algorithmic processes and of sources of data, limiting effective assessment and responsibility. To solve these problems, it is necessary to focus on the long-term policy development, enhance the bargaining power with the vendors, and implement a system of constant ethical analysis.

## VI. TRANSPARENCY AND EXPLAIN ABILITY

#### Transparency in AI Systems

Transparency in library AI systems is the visibility of library discovery tool and metadata generation processes where users are aware of how they are generated (Butt, 2024; Nasr-Azadani and Chatelain, 2024). Metadata transparency tracking provenance means that intuitive results explanations are more crucial to user understanding, whereas on system

openness requires the disclosure of algorithms to provide accountability, and accessibility does not override technical knowledge (Balnaves, 2024).

#### Explainability in AI

Librarians Librarians auditing systems and users verifying recommendations apply this credibility to foster trust and enable detection of bias (Igbinovia et al., 2025). (duration) Metadata provenance logs are used to explain how the NLP model will use certain data, summarizing the decision trees or attention mechanisms as a visualization (e.g., by NLP models), counterfactual explanations (e.g., "this recommendation is higher because of a keyword match X).(duration) (Igbinovia et al., 2025)

#### Vendor Transparency Issues

Vendors such as Ex Libris have proprietary black-box algorithms to find documents that are usually not shared, making them difficult to audit bias, and non-disclosure agreements suppress investigation(Ekstrand et al., 2022; "New Horizons in Artificial Intelligence in Libraries," 2024)

#### Literature Findings

Research studies indicate that there are transparency issues: Explainability is required by OECD and EU AI Act, but on resources, library adoption has hit its limits (Berendt et al., 2023; Nasr-Azadani and Chatelain, 2024). Case studies, such as bias in the rankings of Summon based on skewed corpora, indicate that the vendor is responsible (Berendt et al., 2023).

### VII. ALGORITHMIC BIAS IN DISCOVERY & METADATA

#### Understanding Algorithmic Bias

Algorithms bias happens when the AI results are unfair because of inaccurate data or design (Berendt et al., 2023) Representation bias is the result of underrepresentation of the groups in the training data, which biases such models towards majorities. Historical bias does not eliminate past historical injustices built into data sets, including discriminating categories(Igbinovia and Mensah, 2025). Sampling bias is due to non-representative data collection, which

criticalizes errors in various library collections(Ekstrand et al., 2019).

#### Bias in Library Systems

The AI-based system of automated indexing and classification suggests bias against the works of minority cultures as a result of bilingualism training corpus based on Eurocentric data, misinterpreting, and ignoring other languages and points of view (Igbinovia and Mensah, 2025; Oyighan et al., 2024).

#### Empirical Evidence from Literature

Bias in library AI is reported: Ekstrand et al. indicate discrimination in opaque recommenders, where fairness metrics fail to yield positive results in retrieval(Ekstrand et al., 2019) Wilson makes the case of bias in the discovery layer, such as Summon, which is caused by biased corpora that enforces dominant narratives(Wilson, 2020) Berendt et al. expose a risk of bias in libraries through automated analysis of texts, recommending diversity as a solution to this problem(Berendt et al., 20

#### Consequences of Bias

Bias causes inequitable access with affected and marginalized users having less visibility to useful resources, deepening information divides(Libraries, 2024; Samek and Shiri, 2022). Biased libraries have less chance to garner trust between users and the libraries (Butt, 2024).

### VIII. MITIGATION STRATEGIES AND BEST PRACTICES

The current literature on an ethical use of AI in libraries has found a variety of mitigation measures to deal with the gaps in governance, lack of transparency, and bias in AI-enabled discovery and metadata systems. These approaches may be clustered into five mutually dependent areas namely governance-based measures, transparency and accountability measures, bias reducing measures, capacity building and collaborative model.

#### Governance-Based Approaches

Governance-oriented approaches give more importance to the frameworks that should be established to provide ethical AI implementation in libraries. The research also mentions that internal AI

policies should be in line with more general regulatory and ethical rules, including the EU AI Act and the IEEE Ethically-aligned Design framework, in order to be held accountable, fair, and responsible in use of AI technologies (Balnaves, 2024; Butt, 2024). Moreover, the literature also emphasizes the importance of formal review mechanisms such as AI ethics committees or ethical review boards in the analysis of implementation and effects of AI-supported discovery and metadata systems. Such bodies help governance in terms of human rights as they enable ethical review, risk-taking, and continuous monitoring (Libraries, 2024).

#### Transparency and Accountability Mechanisms

Ethical Book AI is popularly considered to have transparency and accountability as preconditions to its use in libraries. Studies identify the usefulness of explainable AI (XAI) procedures, like Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), or model cards in enhancing AI-based discovery and recommendation methods interpretability (Ekstrand et al., 2022). Tracing and institutional control are made possible through complimentary documentation practices such as the establishment of provenance data on training data and algorithms. It is also evident in the literature that vendors should also be involved to enhance access to audit-relevant information provided by the system, although proprietary platforms can limit this value (Berendt et al., 2023).

#### Bias Reduction Techniques

Mitigation strategies aimed at limiting bias in AI models applied to discover and generate metadata are the identification and reduction of representational, historical, and sampling bias. It is often suggested that NLP and machine learning systems applied to library data should be audited on a regular basis using the algorithm (Berendt et al., 2023). Also, the application of various and non-whites-only datasets is critiqued as one of the methods of combating the dominance of Eurocentric and majority language training corpora (Igbiovvia & Mensah, 2025). Human-in-the-loop can also be used to reinforce ethical decision-making; in this form, the librarians go over such output and evaluate its validity to add professional consideration

and contextual awareness to the automated process (Korošec, 2024).

#### Training and Capacity Building

Capacity building is one of the key elements of ethical implementation of AI in libraries; this trend is established consistently in the literature. Teaching librarians algorithmic literacy allows to consider AI systems responsibly and, thus, make informed decisions. Documentary principles of different organizations, including ALA and IFLA, require continuous education regarding fairness, transparency, and accountability in AI-enabled services (Samek and Shiri, 2022). Professional development programs and workshops based on AI auditing, explainable AI tools, and inclusive data practices have been noted as effective measures to make an institution more prepared and ethical (Steinerova and Ondrisova, 2024).

#### Collaborative Models

Team-based solutions are becoming recognized as the means of dealing with the technical and ethical complexity of AI in libraries settings. Libraries, AI ethicists, computer scientists and system vendors can work together to put together ethically inspired AI pipelines, hybrid explainability medicine, and better evaluation mechanisms (Balnaves, 2024; Berendt et al., 2023). This type of collaboration helps in the shared responsibility and exchange of knowledge building towards stronger and community-sensitive AI practices.

On the whole, each of these mitigation measures is used to form the proactive ethical governance as it promotes transparency and decreases bias, and institutional capacity enhancement. According to the literature, these interrelated practices need to be focused on over a long period to ensure that the implementation of AI in libraries not only does not contradict the principles but also promotes equal access to information (Butt, 2024; Libraries, 2024; Liu and Han, 2024).

## IX. DISCUSSION

#### Synthesis of Key Themes

Governance, transparency, and bias mitigation are the pillars that, identified in this study, that relate to each other to influence ethical AI adoption in library discovery and metadata systems. The regimes linked to the EU AI Act and the Ethically Aligned Design by IEEE focuses on the aspect of accountability based on policy alignment, ethical supervision, periodic auditing of AI systems (Khan et al., 2022; Nasr-Azadani and Chatelain, 2024). By comparison transparency is a existing issue, especially given proprietary vendor-only controlled algorithms employed in discovery systems like Ex Libris and Summon. Lack of transparency of training data and model logic inhibits the ability of the institution to assess the system behavior and detect bias (Ekstrand et al., 2022; Libraries, 2024; New Horizons in Artificial Intelligence in Libraries, 2024).

There are still several types of algorithmic bias, which are representational, historical and sampling bias. These biases affect the relevancy ranking and metadata creation, in most cases favoring the mainstream views that Eurocentric datasets are based on at the expense of allowing many voices to be heard (Berendt et al., 2023; Ekstrand et al., 2019; Igbinoia and Mensah, 2025). These dimensions generate interaction, which is not a one-time affair. Black-box systems with high opaque features may reinforce biased results, including with respect to discovery ranking which strengthen the main stories, whereas an underperforming system of governance constrains institutional reactions to this risk (Balnaves, 2024; Wilson, 2020).

Explainable AI tools, such as LIME and SHAP which serve to enhance the transparency, are highly important protocols that may help balance governance and bias reduction. These tools constitute informed oversight and policy manipulations enabled by the GATTs, which can address audit and interpretability (Ekstrand et al., 2022).

#### Research Gaps Identified

The procedure of review demonstrates that there are a number of gaps in the current scholarship. There is a lack of empirical evidence supporting ethical AI in library-specific situations, especially on the level of

fairness that could be measured through discovery and metadata systems. Standards of explainable artificial intelligence in libraries settings are still developing and no real-world benchmarks can be found particularly in limited resources settings. Moreover, the ongoing lack of transparency in vendors such as non-disclosure agreements and limited API access remain inhibitive factors to fulfill ethical assessment and responsibility.

Implications: Both theoretical and practical.

Policymaking-wise, the results imply the necessity to consider formal governance structures that should be used to facilitate ethical monitoring, such as organization of negotiations with the vendors and creation of ethics monitoring structures and mechanisms based on human rights principles. In the case of system design, the technical and institutional levels of bias can be reduced by integrating hybrid explainable AI and inclusive training datasets. Professionally, ethical implementation of AI systems is possible by enhancing algorithmic literacy through focused training and further education to empower librarians to interact with AI systems more effectively.

Together these insights contribute to a proactive and combined strategy of ethical AI implementation in.

## X. CONCLUSION

### Summary of Findings

This paper outlines some of the ethical issues that revolve around the introduction of artificial intelligence in library discovery systems and metadata systems. The top ones are the bias caused by algorithmic differences when using skewed and mostly Eurocentric datasets, the lack of transparency in proprietary black-box models which conceal operation mechanisms and instability of the functioning controlled by the institutional voids. A variety of mitigation strategies are also found in the literature, such as the introduction of structured governance in place of accountability and auditing, the utilization of explainable AI tools like LIME and SHAP in the name of better interpretability, the creation of inclusive and representative datasets, and

the introduction of human-in-the-loop validation. Enhancing algorithmic literacy in librarians is another method of enforcing ethical control and responsible interaction with the AI systems. Combined, such interrelated methods allow transparency to cause bias detection and governance mechanisms to impose ethical criteria.

#### Recommendations

For library practitioners:

To facilitate equitable discovery results, libraries are encouraged to have regular audits on algorithms, broadcasting hybrid workflows that engage human labor and standard efforts on generating assorted and diverse metadata and maintenance.

For policymakers:

This requires regulatory frameworks requiring more transparency in vendors, as in the EU AI Act, as well as creating ethics review processes and continuing to invest in collaborative AI development, basing such an initiative on fairness and accountability.

For future research:

Future studies ought to address coming up with library-specific measures of fairness, benchmarking increasingly explainable methods and procedures in constrained resources libraries, and exploring approaches to real-time bias monitoring of changing and developing collections.

#### Final Reflections

Ethical AI can empower libraries to become more inclusive and equitable knowledge institutions in aid of equitable representation and accessible knowledge access and informed involved use. Integrating the aspects of fairness, transparency, and accountability into AI-powered systems, libraries will be able to safeguard intellectual freedom and increase diversity of voices in the ever-more AI-mediated information environment. Finally, an ethically and proactive attitude toward AI is guaranteed that technological innovation will increase, but not worsen, the core

library values of accessibility, diversity, and intellectual freedom.

#### REFERENCES

- [1] Abercrombie, G., Benbouzid, D., Giudici, P., Golpayegani, D., Hernández, J., Noro, P., Pandit, H. J., Paraschou, E., Pownall, C., Prajapati, J. M., Sayre, M. A., Sengupta, U., Suriyawongful, A., Thelot, R., Vei, S., & Waltersdorfer, L. (2024). *A collaborative, human-centred taxonomy of AI, algorithmic, and automation harms*. arXiv. <https://doi.org/10.48550/arxiv.2407.01294>
- [2] Balnaves, E. (2024). Artificial intelligence and libraries: An introduction. In *Artificial intelligence in libraries* (pp. 1–20). De Gruyter. <https://doi.org/10.1515/9783111336435-002>
- [3] Berendt, B., Karadeniz, Ö., Kiyak, S., Mertens, S., & d’Haenens, L. (2023). Bias, diversity, and challenges to fairness in classification and automated text analysis: From libraries to AI and back. *arXiv*. <https://doi.org/10.48550/arxiv.2303.07207>
- [4] Butt, J. S. M. (2024). Analytical study of the world’s first EU Artificial Intelligence Act (2024). *International Journal of Research Publication and Reviews*, 5(3), 7343–7352. <https://doi.org/10.55248/gengpi.5.0324.0914>
- [5] Coleman, C. (2020). Managing bias when library collections become data. *International Journal of Librarianship*, 5(1), 8–19. <https://doi.org/10.23974/ijol.2020.vol5.1.162>
- [6] Ekstrand, M. D., Burke, R., & Díaz, F. (2019). Fairness and discrimination in recommendation and retrieval systems. In *Proceedings of the 42nd International ACM SIGIR Conference* (pp. 576–577). ACM. <https://doi.org/10.1145/3298689.3346964>
- [7] Ekstrand, M. D., Das, A., Burke, R., & Díaz, F. (2022). *Fairness in information access systems*. Foundations and Trends® in Information Retrieval, 16(1), 1–177. <https://doi.org/10.1561/15000000079>
- [8] Faife, I. (2020). *IFLA statement on libraries and artificial intelligence*. International Federation of Library Associations and Institutions. <https://repository.ifla.org/handle/123456789/1646> (Accessed March 10, 2025).

- [9] Halburagi, S., & Mukarambi, P. (2023). Use of artificial intelligence technology futures in libraries. *International Journal of Research in Library Science*, 9(2), 1–14. <https://doi.org/10.26761/ijrls.9.2.2023.1642>
- [10] Igbinovia, M. O., & Mensah, M. (2025). Artificial intelligence algorithm bias in information retrieval systems and its implication for library and information science professionals: A scoping review. *Technical Services Quarterly*. Advance online publication. <https://doi.org/10.1080/07317131.2025.2512282>
- [11] Kalisdha, A. (2024). The impact of artificial intelligence and machine learning in library and information science. *International Journal of Research in Library Science*, 10(1), 39–48. <https://doi.org/10.26761/ijrls.10.1.2024.1733>
- [12] Khan, A. A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., Fahmideh, M., Niazi, M., & Akbar, M. A. (2022). Ethics of AI: A systematic literature review of principles and challenges. In *Proceedings of the 2022 ACM/IEEE International Conference* (pp. 383–392). <https://doi.org/10.1145/3530019.3531329>
- [13] Korošec, M. (2024). Developing artificial intelligence in an ethical way in European libraries. In *Artificial intelligence in libraries* (pp. 35–52). De Gruyter. <https://doi.org/10.1515/9783111336435-005>
- [14] Libraries, Association of Research. (2024). *Research libraries guiding principles for artificial intelligence*. <https://doi.org/10.29242/principles.ai2024>
- [15] Liu, L., & Han, M. (2024). Data sharing and exchanging with incentive and optimization: A survey. *Discover Data*, 2(1), Article 6. <https://doi.org/10.1007/s44248-024-00006-2>
- [16] Manjunatha, K. S. (2023). A study on the impact of artificial intelligence on library services. *International Journal of Research in Library Science*, 9(4), 189–198. <https://doi.org/10.26761/ijrls.9.4.2023.1696>
- [17] Nasr-Azadani, M. M., & Chatelain, J. (2024). The journey to trustworthy AI: Part I—Pursuit of pragmatic frameworks. *arXiv*. <https://doi.org/10.48550/arxiv.2403.15457>
- [18] *New horizons in artificial intelligence in libraries*. (2024). De Gruyter. <https://doi.org/10.1515/9783111336435>
- [19] Oyighan, D., Ukubeyinje, E. S., David-West, B. T., & Oladokun, B. D. (2024). The role of AI in transforming metadata management: Challenges, opportunities, and emerging trends. *Asian Journal of Information Science and Technology*, 14(2), 20–29. <https://doi.org/10.70112/ajist-2024.14.2.4277>
- [20] Samek, T., & Shiri, A. (2022). Charting a faceted categorization of AI and ethics. *Proceedings of the Annual Conference of CAIS*. <https://doi.org/10.29173/cais1235>
- [21] Sontan, A. D., & Samuel, S. V. (2024). The intersection of artificial intelligence and cybersecurity: Challenges and opportunities. *World Journal of Advanced Research and Reviews*, 21(2), 1720–1730. <https://doi.org/10.30574/wjarr.2024.21.2.0607>
- [22] Steinerová, J., & Ondrišová, M. (2024). Information ethics in light of bibliometric analyses: Discovering a shift to ethics of artificial intelligence. *Acta Informatica Pragensia*, 13(3), 433–450. <https://doi.org/10.18267/j.aip.237>
- [23] Top trends in academic libraries: A review of the trends and issues. (2022). *College & Research Libraries News*, 83(6), 243–248. <https://doi.org/10.5860/crln.83.6.243>
- [24] Wilson, E. (2020). Masked by trust: Bias in library discovery systems. *Public & Access Services Quarterly*, 16(1), 32–45. <https://doi.org/10.1080/15228959.2019.1697021>