JOURNAL PUSTAKO
Public Universal System for Teaching, Archiving, Knowledge, and Open Access

Algorithmic Bias in AI-Based Library Classification and Recommendation Systems: A Systematic Literature Review

Additya Amanda[1]*
[1]*Universitas Negeri Padang, Indonesia
*Co e-mail: adtyamanda@gmail.com[1]

Article Information
Received: October 20, 2026
Revised: November 09, 2026
Online: November 17, 2026

Keywords
Algorithmic Bias, Artificial Intelligence, Library Classification, Recommender Systems, Algorithmic Fairness, Systematic Literature Review

ABSTRACT
The rapid adoption of artificial intelligence (AI) in library classification and recommendation systems raises concerns about algorithmic bias that may disadvantage certain groups, topics, or materials. This study maps the manifestations and sources of algorithmic bias in AI-based library systems and synthesizes mitigation strategies reported in the literature. A systematic literature review (SLR) was conducted using Scopus, Web of Science, ScienceDirect, DOAJ, and Google Scholar, covering peer-reviewed publications from 1971 to 2026. Based on predefined inclusion and exclusion criteria, 16 sources were selected from an initial pool of more than 80 records and analyzed thematically. The findings identify three interconnected sources of bias: (1) historically biased subject headings and classification structures reused as training data or reference metadata; (2) statistical and popularity biases embedded in machine learning and recommendation algorithms; and (3) limited transparency and auditing mechanisms in commercial discovery systems. Proposed mitigation strategies include systematic bias audits, fairness-aware algorithm design, explainability mechanisms, and improved AI literacy among library and information science (LIS) professionals. This review establishes a conceptual framework connecting critical cataloging scholarship with contemporary algorithmic fairness research. It also highlights the need for empirical investigations into AI bias in library services, particularly in underrepresented contexts, including libraries in developing countries.

DOI: https://doi.org/10.69855/pustako
Available online at website: https://gpijournal.com/index.php/pustako

This work is licensed under a Creative Commons Attribution 4.0 International license

INTRODUCTION

Libraries, including academic, public, and special libraries, are increasingly integrating artificial intelligence (AI) into their core services, ranging from information discovery systems and automated cataloging to reference chatbots and collection recommendation engines. A recent scoping review of 76 studies found that algorithmic bias affects information retrieval systems through biased training data, unequal representation, and limited transparency, raising significant ethical concerns for library services (Igbinovia & Danquah, 2025). Similarly, Cox (2022) argues that AI adoption in information services introduces ethical challenges that information professionals must address, particularly regarding data privacy, transparency, and accountability for algorithmic decisions. These developments indicate that evaluating AI in libraries requires attention not only to operational efficiency but also to the fairness and inclusiveness of the information provided to users.

Algorithmic bias generally refers to the systematic tendency of an AI system to produce outputs that disproportionately favor or disadvantage particular groups, topics, or entities. Such bias may originate from unrepresentative training data, model design choices, or algorithmic decision-making processes (Mehrabi et al., 2021). Kordzadeh and Ghasemaghaei (2022) further explain that algorithmic bias is not exclusively a technical problem but a sociotechnical phenomenon shaped by the interaction between system design, its implementation context, and broader social structures. In library environments, this perspective is particularly important because AI systems operate on knowledge structures and metadata that are themselves products of historical, cultural, and institutional decisions.

Bias in library information organization predates the development of AI. Critical librarianship has long examined how classification schemes and subject headings can reproduce unequal representation. Berman (1971), through *Prejudices and Antipathies*, documented numerous Library of Congress Subject Headings (LCSH) that contained discriminatory or inappropriate terminology concerning racial, ethnic, and gender groups. Knowlton (2005) subsequently found that although many of Berman's proposed revisions had been adopted by the Library of Congress, several problematic headings persisted. Similarly, Olson (2001, 2002), drawing on feminist and poststructuralist perspectives, demonstrated how the hierarchical and universalist structures of library classification systems can marginalize topics concerning women, minority groups, and non-Western knowledge. Olson and Schlegl (1999) extended this discussion through a content analysis of 93 works of critical literature, identifying biases related to gender, sexuality, race, age, disability, ethnicity, language, and religion in subject access standards. Martínez-Ávila and Guimarães (2013) also argued that claims of neutrality and universality in classification systems, including the Dewey Decimal Classification and Library of Congress Classification, may conceal underlying cultural and ethical assumptions.

The relationship between historical classification bias and contemporary AI introduces an additional challenge for libraries. When existing classification structures, subject headings, and bibliographic metadata are incorporated into AI-based systems, their embedded assumptions may be transferred into automated classification, intelligent information retrieval, and collection recommendation processes. Reidsma documented instances of indexical bias in the Topic Explorer feature of the Summon discovery system, in which certain search terms generated results

This work is licensed under a Creative Commons Attribution 4.0 International license

associated with discriminatory or socially sensitive meanings. Similarly, Auraria Library acknowledged the potential for algorithmic bias in the relevance ranking, recommendation, and facet-filtering functions of its commercial discovery system, leading to the establishment of a user feedback mechanism for reporting such problems (American Library Association [ALA], 2021). These examples demonstrate that algorithmic bias is not limited to the internal operation of AI models but can also influence how library resources are represented, ranked, and accessed by users.

Research on recommender systems in computer science further demonstrates that algorithmic bias can emerge through several mechanisms. One prominent example is popularity bias, in which recommendation algorithms repeatedly prioritize already popular items while reducing the visibility of less popular materials in the collection's long tail (Carnovalini et al., 2025). Wang et al. (2023), through a systematic review of more than 60 studies, identified bias at the data, model, and output levels, with implications for both process fairness and outcome fairness. Deldjoo et al. (2023) further highlighted that fairness research in recommender systems does not always explicitly define the normative principles or disadvantaged groups underlying its fairness evaluations. Meanwhile, Orphanou et al. (2022) proposed a cross-domain framework for addressing algorithmic bias through three interconnected stages: bias detection, fairness management, and explainability management. Although these studies provide important foundations for understanding algorithmic fairness, their findings need to be considered alongside the specific characteristics of library knowledge organization, where representational bias in metadata and classification systems may influence subsequent algorithmic outputs.

Despite the development of critical cataloging scholarship and contemporary research on algorithmic fairness, these two research traditions have largely been examined separately. Critical cataloging studies primarily investigate representational and cultural bias in traditional knowledge organization, whereas AI fairness studies generally emphasize data, model, and algorithmic performance. Consequently, the relationship between historically embedded bias in library knowledge structures and its potential reproduction or amplification through AI-based classification and recommendation systems remains insufficiently synthesized. This gap is particularly relevant to libraries in developing countries such as Indonesia, where the adoption of AI-based services is emerging but may be accompanied by limitations in AI literacy, institutional evaluation capacity, and guidelines for assessing algorithmic fairness. Furthermore, AI tools developed using predominantly Western and English-language data, as well as established classification structures such as DDC and LCSH, may not adequately represent local languages, cultural contexts, and indigenous knowledge. Therefore, a systematic synthesis is necessary to connect historical perspectives on library classification bias with contemporary approaches to algorithmic fairness and to identify implications for more inclusive library services.

The novelty of this study lies in its integration of critical cataloging scholarship and contemporary algorithmic fairness research into a consolidated conceptual framework for understanding bias in AI-based library classification and recommendation systems. Rather than treating historical classification bias and technical algorithmic bias as separate problems, the proposed framework conceptualizes them as interconnected stages of a sociotechnical process. It links (1) historical knowledge structures and subject-heading bias, (2) the potential transfer and

This work is licensed under a Creative Commons Attribution 4.0 International license

amplification of bias through AI models and recommendation algorithms, (3) the manifestation of bias in information discovery, resource ranking, and collection recommendations, and (4) interventions involving bias auditing, fairness management, explainability, and institutional policies. This integration provides a broader analytical perspective for examining how bias originates, develops, becomes visible in library services, and can be addressed.

Accordingly, this study aims to: (1) identify the sources and manifestations of algorithmic bias in AI-based library classification and recommendation systems; (2) integrate findings from critical cataloging literature and contemporary algorithmic fairness research; and (3) identify strategies for mitigating algorithmic bias in library practice. The study is guided by three research questions:

- RQ1: What are the sources of algorithmic bias in AI-based library classification and recommendation systems?
- RQ2: How does algorithmic bias manifest in library classification, information discovery, and recommendation services?
- RQ3: What mitigation strategies have been proposed in the literature to reduce the impact of algorithmic bias in library systems?

The contribution of this study is both conceptual and practical. Conceptually, it provides an integrated framework that connects historical representational bias with contemporary algorithmic processes, extending the discussion of fairness beyond technical model performance to include the knowledge structures underlying library systems. Practically, the framework offers a basis for librarians, information professionals, and library institutions to evaluate metadata quality, identify potential bias in AI-generated outputs, strengthen transparency and accountability, and develop more inclusive digital library services.

METHODS

This study employs a systematic literature review (SLR) approach, guided by the general principles of transparent review reporting namely identification, screening, eligibility assessment, and inclusion of sources.

1. Data Sources and Search Strategy

Literature searches were conducted across major academic databases, namely Scopus, Web of Science, ScienceDirect, the Directory of Open Access Journals (DOAJ), the ACM Digital Library, and Google Scholar as a supplementary source to capture relevant institutional repositories and preprints. The search strategy combined the following keywords and Boolean operators: ("algorithmic bias" OR "algorithm bias" OR "AI bias") AND ("library" OR "classification" OR "cataloging" OR "subject heading") AND ("recommender system" OR "recommendation" OR "discovery system" OR "information retrieval"). The search was further expanded with related terms such as "fairness in recommender systems," "bias in library classification," and "critical cataloging" to capture the classic literature on bias in library classification schemes that forms the historical root of this issue.

2. Inclusion and Exclusion Criteria

The inclusion criteria applied were: (1) journal articles, conference proceedings, or book chapters that had undergone peer review, or seminal works widely cited within the critical-

This work is licensed under a Creative Commons Attribution 4.0 International license

cataloging field; (2) publications dated between 1971 and 2026, in order to capture both the historical roots of classification critique and recent developments in AI algorithmic bias; (3) explicit discussion of bias, fairness, or ethics within the context of library classification, discovery systems, or algorithm/AI-based recommendation systems; and (4) availability in English or Indonesian with full-text access.

The exclusion criteria comprised: (1) articles that discussed AI or library systems in purely technical terms without any connection to bias, fairness, or ethical issues; (2) opinion pieces, blog posts, or editorial notes lacking a systematic literature basis; and (3) duplicate records identified across multiple databases.

3. Selection Process and Analysis

The To ensure a transparent and rigorous selection process, the literature screening followed a multi-stage adaptation of the PRISMA guidelines. The initial database search yielded a total of 85 records. After removing 15 duplicate records, 70 unique records proceeded to the title and abstract screening phase. During this phase, 42 records were excluded because they did not meet the core thematic scope (e.g., purely technical AI papers without library context, or general library studies unrelated to bias).

The remaining 28 full-text articles were then assessed for eligibility against all inclusion criteria. Out of these, 12 full-text articles were further excluded due to specific reasons: 5 lacked sufficient academic depth or peer-review validation, 4 focused exclusively on traditional cataloging without addressing modern digital/AI shifts or systematic evaluation, and 3 were inaccessible in full-text English or Indonesian. Ultimately, 16 sources met all criteria and were included for qualitative thematic synthesis in this review.

The selection breakdown across these stages is summarized in Table 1 below.

Systematic Review Stage	Number of Records / Sources
Identification: Initial records retrieved from databases (Scopus, Web of Science, ScienceDirect, DOAJ, ACM Digital Library, and Google Scholar)	85
Screening 1: Duplicate records removed	15
Screening 2: Records screened by title and abstract	70
Screening 3: Records excluded based on title and abstract screening	42
Eligibility: Full-text articles assessed for eligibility	28
Eligibility Exclusion: Full-text articles excluded due to insufficient academic rigor, off-topic focus, or inaccessible full text	12
Included: Sources included in the qualitative thematic synthesis	16

RESULTS

Thematic analysis of the 16 selected sources produced five main themes, which are consolidated into an integrated framework: (1) the historical roots of bias in library classification systems; (2) the sources and characteristics of algorithmic bias in AI systems in general; (3) the manifestation of bias in library discovery services and AI applications; (4) bias in collection recommender systems; and (5) mitigation strategies identified in the literature. These themes and their interconnections are elaborated below, preceded by a conceptual synthesis framework.

This work is licensed under a Creative Commons Attribution 4.0 International license

Table 2. Summary of the 16 Selected Sources

No.	Author(s)	Year	Main Focus
1	Berman	1971	Bias in LCSH
2	Olson & Schlegl	1999	Bias in subject access standards
3	Olson	2001	Representation in library catalogs
4	Olson	2002	Subject representation
5	Knowlton	2005	Changes in LCSH
6	Martínez-Ávila & Guimarães	2013	Library classification criticism
7	Mehrabi et al.	2021	Bias and fairness in machine learning
8	ALA	2021	Algorithmic bias in library systems
9	Cox	2022	Ethics of AI for information professionals
10	Kordzadeh & Ghasemaghaei	2022	Algorithmic bias
11	Orphanou et al.	2022	Bias mitigation
12	Deldjoo et al.	2023	Fairness in recommender systems
13	Wang et al.	2023	Recommender-system fairness
14	Carnovalini et al.	2025	Popularity bias
15	Igbinovia & Danquah	2025	AI bias in information retrieval
16	Reidsma	n.d.	Algorithmic bias in discovery systems

Source: Authors' synthesis of the selected literature.

To explicitly address the linkage between classical critical cataloging and modern AI ethics, this review proposes a conceptual synthesis framework. The framework visualizes a cascading socio-technical pipeline wherein algorithmic bias in modern library AI systems is not an isolated technical bug, but a multi-layered phenomenon originating from deep-rooted historical structures. Specifically, the framework illustrates that:

- **Layer 1 (Root/Foundation):** Historical knowledge structures and subject headings (e.g., Berman, 1971; Olson, 2002) form the foundational metadata and training data used in modern digital libraries.
- **Layer 2 (Technical Amplification):** When these historical structures feed into machine learning models and discovery/recommender systems, they interact with technical flaws such as popularity bias, representation bias, and optimization choices (Mehrabi et al., 2021; Wang et al., 2023; Carnovalini et al., 2025).
- **Layer 3 (Interface Manifestation):** Combined with a lack of transparency in commercial systems (Igbinovia & Danquah, 2025), this pipeline manifests at the user interface as distorted search results, biased rankings, and long-tail marginalization (Reidsma; American Library Association, 2021).
- **Layer 4 (Intervention/Mitigation):** Counteracting this requires a synchronized response spanning technical audits, explainability management, data curation, and institutional LIS policies (Orphanou et al., 2022; Igbinovia & Danquah, 2025).

1. Historical Roots of Bias in Library Classification Systems

Prior to the AI era, bias in libraries had long been examined through critical approaches to classification schemes and subject headings. Berman (1971) stands as the seminal reference that systematically documented 225 LCSH terms judged to be biased against particular racial, ethnic,

This work is licensed under a Creative Commons Attribution 4.0 International license

religious, and other groups, while also proposing revisions to those terms. Knowlton (2005), tracing developments over the following three decades, found that although many of Berman's proposals had been adopted, several areas of bias persisted, indicating that changes to formal knowledge structures move far more slowly than shifts in social norms. Olson (2001, 2002) deepened this critique through feminist and poststructuralist lenses, showing that the principle of universality underpinning classification schemes such as the Library of Congress Classification (LCC) produces hierarchical and binary relations that structurally marginalize the representation of women and non-dominant groups. Olson and Schlegl (1999) broadened the scope of this critique empirically, through a content analysis of 93 works of literature mapping bias by gender, race, age, disability, ethnicity, language, and religion in subject access standards. Martínez-Ávila and Guimarães (2013) further argued that claims of neutrality in classification schemes contradict the reality that every classification scheme is a cultural product of its creators, and therefore inherently carries certain epistemic biases.

2. Sources and Characteristics of Algorithmic Bias in AI Systems

At a more technical level, Mehrabi et al. (2021) categorized the sources of bias in machine learning systems into three broad groups data bias, algorithmic bias, and societal bias which interact with one another throughout the AI development lifecycle, from data collection and model building to deployment and user interaction. Kordzadeh and Ghasemaghaei (2022) complement this framework by emphasizing that algorithmic bias cannot be understood merely as a technical error, but rather as a socio-technical phenomenon involving design choices, context of deployment, and broader social power structures. Orphanou et al. (2022), through a cross-domain review spanning machine learning, recommender systems, information retrieval, and human–computer interaction research, identified that a comprehensive treatment of algorithmic bias requires three interrelated stages bias detection, fairness management, and explainability management involving multiple stakeholders ranging from developers and regulators to both direct and indirect users of the system.

3. Manifestation of Bias in Library Discovery Services and AI Applications

Within the direct context of library services, Igbinovia and Danquah (2025), through a scoping review of 76 studies, found that AI algorithmic bias affects library information retrieval systems through three main pathways: biased training data, disproportionate group representation, and a lack of algorithmic transparency, which together give rise to ethical concerns regarding the credibility and fairness of the system. Concrete case documentation was provided by Reidsma, who identified patterns of indexical bias in the Topic Explorer feature of the Summon discovery system, where searches on particular themes were directed toward results carrying discriminatory or socially insensitive connotations. These field-level findings were further corroborated by Auraria Library, which openly acknowledged that the commercial discovery product it licenses carries the potential for bias in relevance ranking, facet filtering, and resource recommendation, prompting it to build a user-feedback mechanism to identify and report such bias to the vendor (American Library Association, 2021). From an ethical standpoint, Cox (2022) outlined eight ethical scenarios facing information professionals in the deployment of AI in

This work is licensed under a Creative Commons Attribution 4.0 International license

libraries, spanning issues of user data privacy, accountability for algorithmic decisions, and the potential for bias in automated recommendation and classification systems.

4. **Bias in Collection Recommender Systems**

On the more specific matter of recommender systems, Wang et al. (2023), through a systematic review of more than 60 studies across leading conferences and journals, showed that bias in recommender systems can arise at the level of data, model, or output, and produces two forms of unfairness process unfairness and outcome unfairness. Carnovalini et al. (2025) specifically highlight the phenomenon of popularity bias, whereby recommendation algorithms persistently favor already-popular items, making items in the collection's "long tail" including special collections or niche reading interests increasingly difficult for users to discover even when relevant to their needs. Deldjoo et al. (2023) add a critical note that most fairness research in recommender systems has yet to explicitly articulate the normative fairness claim being addressed, or which groups are considered disadvantaged, leaving room for implicit bias in the design of fairness metrics themselves. These three findings bear direct relevance to collection recommender systems in digital libraries, which risk replicating similar bias patterns if designed without accounting for diversity and balance within the collection.

5. **Mitigation Strategies Identified in the Literature**

The reviewed literature proposes a range of mitigation strategies that can be grouped at the technical level and the institutional/professional level. At the technical level, Orphanou et al. (2022) propose a three-stage framework: bias detection through systematic audits of algorithmic outputs, fairness management through debiasing techniques applied to data and models, and explainability management so that algorithmic decisions can be understood and held accountable by stakeholders. At the institutional and professional level, Igbinovia and Danquah (2025) emphasize the strategic role of LIS professionals in mitigating bias through five pathways: strengthening information literacy and AI literacy among users, conducting regular algorithmic audits of licensed systems, more ethical data curation, collaboration with technology vendors, and policy advocacy at both the institutional and national levels. In a similar vein, Auraria Library practices a participatory approach by opening a bias-reporting channel for users as a form of collective audit, subsequently relayed to the discovery-system vendor (American Library Association, 2021). At the root-cause level, efforts to revise biased subject headings, as pioneered by Berman (1971) and continued by the critical cataloging movement (Knowlton, 2005; Olson, 2002), remain relevant as a long-term foundation, given that biased data and knowledge structures risk being continuously inherited by AI systems if left uncorrected at the source.

DISCUSSION

The synthesis of the five themes indicates that algorithmic bias in AI-based library classification and recommendation systems is not an entirely new phenomenon but rather a continuation and potential amplification of representational inequalities historically identified in library knowledge organization. This finding supports the critical-cataloging arguments of Berman (1971), Olson (2001, 2002), and Olson and Schlegl (1999), who demonstrated that classification

This work is licensed under a Creative Commons Attribution 4.0 International license

schemes and subject headings may reproduce cultural, gender, racial, and social biases. The present review extends these arguments by highlighting how such historical biases may be transferred into contemporary AI systems when existing classification structures and bibliographic metadata are used as reference data or training resources. Thus, bias in library AI systems should be understood not only as a consequence of technical limitations but also as a problem rooted in the historical organization and representation of knowledge.

Regarding RQ1, the findings identify three interconnected sources of algorithmic bias: historical representational bias, technical bias, and institutional limitations in transparency and accountability. The first source is consistent with Olson's (2002) argument that classification systems are not neutral representations of knowledge because they reflect particular cultural and ideological perspectives. Similarly, Knowlton (2005) demonstrated that revisions to biased Library of Congress Subject Headings did not eliminate all problematic representations. These findings suggest that improving AI-based classification requires attention to the quality and inclusiveness of the underlying metadata, rather than relying exclusively on technical improvements to the algorithm. At the technical level, Mehrabi et al. (2021) explain that bias can emerge from training data, model design, and broader societal conditions. Wang et al. (2023) further identify data-, model-, and output-level bias in recommender systems. The present review connects these perspectives by showing how historical representational bias may interact with technical mechanisms, potentially reinforcing unequal visibility within library information services. In addition, Igbinovia and Danquah (2025) and the American Library Association (2021) highlight the importance of transparency and institutional accountability, particularly when libraries depend on commercial discovery platforms whose internal ranking and recommendation processes may not be fully accessible.

Regarding RQ2, the findings demonstrate that algorithmic bias can become visible through distorted search associations, unequal resource representation, and recommendation patterns that favor popular materials over less visible collection items. This interpretation is consistent with Reidsma's documentation of indexical bias in the Summon Topic Explorer feature, where particular search terms were associated with socially sensitive or discriminatory meanings. Such cases illustrate that bias can influence not only which resources are retrieved but also how subjects and communities are represented through search interfaces. In recommender systems, the findings correspond with Carnovalini et al. (2025), who discuss popularity bias and the reduced visibility of long-tail items. Wang et al. (2023) provide a broader explanation by distinguishing between process fairness and outcome fairness, suggesting that apparently efficient recommendation processes may still produce unequal outcomes. Deldjoo et al. (2023) further emphasize that fairness assessments must explicitly identify the normative principles and disadvantaged groups being considered. Taken together, these studies indicate that evaluating library recommendation systems solely through relevance or user engagement metrics may overlook broader questions of representational diversity and equitable access to information. This is particularly important for collections containing local knowledge, minority perspectives, and specialized academic materials that may receive less algorithmic visibility.

Regarding RQ3, the review indicates that mitigating algorithmic bias requires coordinated technical, institutional, and professional interventions. At the technical level, Orphanou et al.

This work is licensed under a Creative Commons Attribution 4.0 International license

(2022) propose an integrated approach involving bias detection, fairness management, and explainability management. This framework is relevant to library systems because identifying biased outputs alone is insufficient without subsequent corrective action and mechanisms for explaining algorithmic decisions. The present review extends this approach by emphasizing the need to address the historical and representational quality of library metadata as an additional consideration in bias prevention. At the institutional level, Igbinovia and Danquah (2025) emphasize the role of library and information science professionals in promoting ethical AI practices, strengthening AI literacy, and evaluating algorithmic systems. Cox (2022) similarly highlights the ethical responsibilities of information professionals in addressing accountability and transparency in AI-supported services. These arguments support the position that librarians should not be treated merely as end users of automated systems but as active participants in evaluating their fairness and social implications. Practical interventions may include reviewing subject headings and metadata, conducting periodic audits of search and recommendation outputs, requesting greater transparency from technology vendors, and establishing channels through which users can report potentially biased results.

The conceptual contribution of this review is the integration of historical classification bias and contemporary algorithmic fairness into a four-layer framework. The first layer concerns historical knowledge structures, including subject headings and classification schemes that may contain representational inequalities. The second layer concerns technical amplification, in which existing biases may interact with training data, machine-learning models, and recommendation algorithms. The third layer concerns interface manifestation, where bias may become visible through search rankings, resource recommendations, and unequal information visibility. The fourth layer concerns intervention and mitigation, involving metadata review, algorithmic auditing, fairness management, explainability, and institutional policies. Unlike approaches that primarily examine bias at the data, model, or output levels, this framework emphasizes the relationship between the historical organization of knowledge and the subsequent technical and institutional processes through which bias may be reproduced. Its contribution is therefore to provide an integrated analytical perspective for understanding the possible progression of bias from knowledge structures to user-facing library services. However, this framework is a conceptual synthesis derived from the reviewed literature and requires empirical validation in operational library environments.

The practical implications are particularly relevant to libraries in developing countries, including Indonesia. The reviewed literature indicates that AI-based library services may face challenges related to the representation of local languages, cultural knowledge, and underrepresented collections. Although the selected literature provides limited direct empirical evidence from Indonesian libraries, the synthesized framework offers several potential directions for institutional practice. First, libraries can conduct collaborative audits of bibliographic metadata, subject headings, and repository records to identify potentially problematic or unrepresentative descriptions. Second, institutions can incorporate transparency, accountability, and bias-reporting requirements into the procurement and evaluation of commercial discovery systems. Third, explainability guidelines can be incorporated into digital library workflows to support the evaluation of automated classification and recommendation outputs. Fourth, targeted AI literacy

This work is licensed under a Creative Commons Attribution 4.0 International license

programs can help librarians and users recognize potential bias and understand the limitations of algorithmically generated search results. These measures should be adapted to institutional resources and local linguistic and cultural contexts rather than applied as a uniform technical solution.

Several limitations should also be considered when interpreting the findings. First, the reviewed literature is predominantly drawn from North American and European contexts, limiting the extent to which the findings can be generalized to libraries in developing countries, including Indonesia. Second, many of the selected sources are conceptual or literature-based, while empirical studies directly auditing AI-based classification and recommendation systems in operational libraries remain limited. Third, the review included only English- and Indonesian-language sources, potentially excluding relevant research published in other languages. Furthermore, because the review integrates foundational critical-cataloging scholarship with contemporary AI fairness research, the selected literature covers different technological periods and research contexts. Therefore, the proposed relationships between historical classification bias and contemporary algorithmic amplification should be interpreted as a conceptual synthesis rather than as a demonstrated causal relationship. Future studies should conduct empirical audits of library discovery systems, examine the representation of local and indigenous knowledge in digital collections, and evaluate the effectiveness of proposed mitigation strategies across different institutional and cultural settings.

CONCLUSIONS

This systematic literature review concludes that algorithmic bias in library AI systems is a multi-layered socio-technical challenge driven by historical cataloging structures, machine learning statistical flaws, and proprietary opacity. To address these challenges, libraries must adopt a categorized tripartite framework:

- Technical strategies: implementing bias-detection audits, algorithmic debiasing, and explainability management.
- Institutional strategies: establishing vendor-evaluation standards, creating ethical AI guidelines, and developing institutional bias-audit policies for discovery systems.
- Professional strategies: providing targeted AI literacy training and driving collaborative data curation among librarians.

The primary contribution of this work bridges classical critical-cataloging scholarship with contemporary algorithmic fairness research. For developing contexts such as Indonesia, future initiatives must operationalize these frameworks by formulating national or institutional procurement standards, conducting empirical field audits on local repositories, and preparing LIS professionals to proactively manage AI-driven library services.

REFERENCES

American Library Association, Office for Diversity, Literacy, and Outreach Services. (2021). Auraria Library: Taking steps to address algorithmic bias in library systems. https://www.ala.org/advocacy/diversity/odlos-blog/algorithmic-bias-libraries

Berman, S. (1971). Prejudices and antipathies: A tract on the LC subject heads concerning people. Scarecrow Press.

This work is licensed under a Creative Commons Attribution 4.0 International license

Carnovalini, F., Rodà, A., & Wiggins, G. A. (2025). Popularity bias in recommender systems: The search for fairness in the long tail. Information, 16(2), Article 151. https://doi.org/10.3390/info16020151

Cox, A. (2022). The ethics of AI for information professionals: Eight scenarios. Journal of the Australian Library and Information Association, 71(3), 201–214. https://doi.org/10.1080/24750158.2022.2084885

Deldjoo, Y., Jannach, D., Bellogín, A., Difonzo, A., & Zanzonelli, D. (2024). Fairness in recommender systems: Research landscape and future directions. User Modeling and User-Adapted Interaction, 34, 59–108. https://doi.org/10.1007/s11257-023-09364-z

Igbinovia, M. O., & Danquah, M. 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, 42(3–4), 253–279. https://doi.org/10.1080/07317131.2025.2512282

Knowlton, S. A. (2005). Three decades since Prejudices and Antipathies: A study of changes in the Library of Congress Subject Headings. Cataloging & Classification Quarterly, 40(2), 123–145. https://doi.org/10.1300/J104v40n02_08

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. European Journal of Information Systems, 31(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Martínez-Ávila, D., & Guimarães, J. A. C. (2013). Library classifications criticisms: Universality, poststructuralism and ethics. Scire, 19(2), 21–26. http://ibersid.eu/ojs/index.php/scire/article/view/4081

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), Article 115. https://doi.org/10.1145/3457607

Olson, H. A. (2001). The power to name: Representation in library catalogs. Signs: Journal of Women in Culture and Society, 26(3), 639–668. https://doi.org/10.1086/495624

Olson, H. A. (2002). The power to name: Locating the limits of subject representation in libraries. Kluwer Academic Publishers.

Olson, H. A., & Schlegl, R. (1999). Bias in subject access standards: A content analysis of the critical literature. Proceedings of the Annual Conference of the Canadian Association for Information Science/Actes du congrès annuel de l'Association canadienne des sciences de l'information. https://doi.org/10.29173/cais460

Orphanou, K., Otterbacher, J., Kleanthous, S., Batsuren, K., Giunchiglia, F., Bogina, V., Shulner-Tal, A., Hartman, A., & Kuflik, T. (2023). Mitigating bias in algorithmic systems—A fish-eye view. ACM Computing Surveys, 55(5), Article 87. https://doi.org/10.1145/3527152

Reidsma, M. (2016, March 11). Algorithmic bias in library discovery systems. Matthew Reidsma. https://matthew.reidsrow.com/articles/173

Wang, Y., Ma, W., Zhang, M., Liu, Y., & Ma, S. (2023). A survey on the fairness of recommender systems. ACM Transactions on Information Systems, 41(3), Article 52. https://doi.org/10.1145/3547333