Majestic Multidisciplinary Science Insight
ISSN: 3142-8436
Volume 1, 2026, **pp.** 1-18, **Article ID:** CM2601107004
DOI: 10.65080/msi.v1.CM2601107004

Systematic Review Article

Generative AI in Academic Libraries: A Systematic Review of AI-Mediated Reference Services, Librarian Roles, and User Trust

Arslan Sheikh[1, *]

[1]*COMSATS University Islamabad, Islamabad Campus, Islamabad, Pakistan*

Article History
Received: 22 February, 2026 Revised: 30 April, 2026 Accepted: 03 June, 2026 Published: 21 September, 2026

Abstract:

Introduction: Generative Artificial Intelligence (AI), and in particular large language models such as ChatGPT, is being gradually integrated into the systems and workflows of academic libraries. Consequently, it is transforming the nature of reference services, reshaping the work of librarians, and influencing the information-seeking behaviour of users. This systematic review aimed to explore the impact of generative AI on the quality of reference services, restructuring of librarian tasks, and trust of users in academic libraries.

Methodology: A systematic search of Scopus and ScienceDirect was conducted to identify relevant literature. A total of 13 major empirical studies, published between 2023 and 2026 were selected and analysed that included experimental assessments, cross-sectional surveys, qualitative interviews, mixed-methods research and quasi-experimental audit studies across different geographic locations.

Results: The findings revealed a significant performance of AI-mediated reference services but generative AI can only be applied to low-complexity questions. However, it demonstrated severe limitations in responding to complex research questions, contextual questions, and citation verification, as indicated by high rates of hallucinations and bibliographic errors. However, user trust was also conditional, depending on the perceived utility, transparency, and the institutionalisation of AI services.

Conclusion: Generative AI should be understood as a socio-technical system, in which the efficacy is determined by human control, governance systems, and correlation with the fundamental values, which are the foundation of academic libraries. Thus, this study emphasises that, instead of replacing librarians, generative AI restructures professional practices in favour of mediation, verification, ethical oversight, and the delivery of AI literacy education.

Keywords: Generative artificial intelligence, academic libraries, reference services, librarian roles, user trust, information integrity, ChatGPT.

1. INTRODUCTION

Chatbots based on Generative Artificial Intelligence (GenAI), specifically, Large Language Models (LLMs), are swiftly transforming academic communities. They help users to discover information, evaluate evidence, and obtain help with conducting research and composing scholarly writing. Student uptake is now high enough to affect the demand profile for academic library services. The UK Higher Education Policy Institute (HEPI, 2025) Student Generative AI Survey 2025 reports that 88% of students reported using generative AI tools for assignments (up from 53% the previous year). Although GenAI is becoming part of the daily routines of academic research, academic libraries are under ever-increasing pressure to act across three interconnected areas: maintaining the quality of reference services, recalibrating labour roles and skills, and preserving users' trust. Early experimental findings indicate that GenAI can offer daily advice but not the highest-quality reference results. In chronologically organised testing of ChatGPT across different types and complexity levels of reference queries, (Lai, 2023) found that the system performed poorly on advanced research questions and complex inquiries, including precisely the kinds of queries that characterise academic library reference work. One of the main weaknesses

*Address correspondence to this author at COMSATS University Islamabad, Islamabad Campus, Islamabad, Pakistan; E-mail: arslan_sheikh@comsats.edu.pk

Cite as: Sheikh, A. (2026). Generative AI in academic libraries: A systematic review of AI-mediated reference services, librarian roles, and user trust. *Majestic Multidisciplinary Science Insight*, *1*, 1-18. Article ID: CM2601107004. https://doi.org/10.65080/msi.v1.CM2601107004

is the model's tendency to produce plausible but incorrect statements, such as bibliographic errors. A socio-technical perspective in library practice acknowledges that AI results are governed by organisational settings, staff processes, professionalism, and policy alignment within an institution. In the cross-disciplinary study by (Mugaanyi *et al.,* 2024), approximately 60% of the recalled outputs contained correct citations. In the case of libraries, these findings indicate a risk that, provided the use of generative AI to answer reference queries, recommend sources, or teach citation practice, there is a possibility of such uncorrected inaccuracies. These inaccuracies may lead to poor service provision and loss of confidence in library directions as the users make mistakes.

Simultaneously, there is also more effort to apply generative AI to the organisational level. According to the evidence presented by academic libraries, generative AI is being implemented in service delivery, user care, and internal processes. Indicatively, (Gmiterek & Kotuła, 2025) used a survey and content analysis to identify the degree of generative-AI adoption by academic libraries at Polish state universities (a mixed-methods study). The workforce implication is also evident, as (Cox *et al.,* 2019) UK-wide survey of AI and the library profession found that the current use and perception of AI by librarians and information professionals are now regarded as a major professional concern rather than being an issue of the future. User trust is often underestimated but is the third variable indicating the connection between the quality of references and the workforce change. As (Deschênes & McMahon, 2024) reported on the use of generative-AI chatbots in academic research, students use these tools but, according to the survey, they do not consider the information produced by generative AI helpful. As (Grams, 2024) notes, perceived usability advantages are in a better position than human reference interactions. However, the comfort factor will also tend to affect trust and preference where there are problems of credibility.

Although there is an increased number of empirical studies of generative AI in academic libraries, the research remains sporadic, which is why a specific systematic review is necessary. The literature at hand analyses the independent factors of AI use. It, however, does not centre on the joint impact on library services. Indicatively, the correctness of answers and validity of references are the major measures of reference quality as applied by (Lai, 2023). However, these outcomes are not explicitly related to these measures of performance that are so critical to workforce redesign like models of supervision, timely mediation, workflows of quality-assurance mechanisms, or objective markers concerning user trust. (Cox *et al.,* 2019) concentrated on professional expectations and views under a sample survey of librarians, however, the study lacked comprehensive data on workforce perception, measured performance of services, and patron confidence in academic libraries. (Deschênes & McMahon, 2024) documented patterns of use and attitudes towards generative AI but did not compare differences in trust between AI use within library-branded services and independent use of external AI tools by students.

The following review tackles these gaps by synthesising the interdependencies between the performance of AI in academic libraries, work and mediation in the field of libraries, and the role of trust in the formation of relationships. The review does not consider reference quality, workforce transformation and user trust as independent components, but rather sees them as mutually reinforcing socio-technical processes which influence the integration of generative AI into academic library services. This systematic review aims to summarise recent empirical evidence on the effects of generative AI on academic library services and professional practice. Notably, it explores implications on the quality of reference services such as accuracy, completeness, contextual relevance, and citation integrity, changes in librarian roles, competencies, and trends in user trust. The review will help illuminate evidence-based governance and the responsible incorporation of generative AI in academic libraries by providing comprehensive evidence.

1.1. Conceptual Framework

Drawing on these three factors, this review uses a socio-technical conceptual framework to describe how generative AI can be integrated into academic libraries, as the three dimensions interdependently influence one another (Table **1**). The framework builds on the assumption that the quality of interactions with AI-generated responses such as accuracy of citations, relevance to context and reliability of information influence users' confidence in AI-supported library services. As inaccuracies, hallucinations, or trust issues arise, librarians play greater roles in supervision and governance in validating research, ensuring ethical management and AI literacy support. Instead, successful workforce mediation could foster trust in the institutions and provide a means to adopt AI responsibly. For this reason, generative AI is explored as part of a socio-technical system where service quality, professional labour, and trust continuously shape one another. This framework also led to the synthesis and categorisation of the studies included in the review (Fig. **1**).

Table 1. Mapping of studies to the conceptual framework.

Framework Dimension	Representative Studies
AI performance & reference quality	(Lai, 2023; Mugaanyi *et al.,* 2024; Grams, 2024)
User trust & adoption behaviour	(Grams, 2024; Deschênes & McMahon, 2024; Haris *et al.,* 2025; Khan *et al.,* 2026)
Workforce mediation & governance	(Chen, 2026; Gmiterek & Kotuła, 2025; Khan, 2025; Huang *et al.,* 2023; Khan *et al.,* 2026)
Interdependent socio-technical dynamics	(Khan, 2025; Huang *et al.,* 2023; Khan *et al.,* 2026)

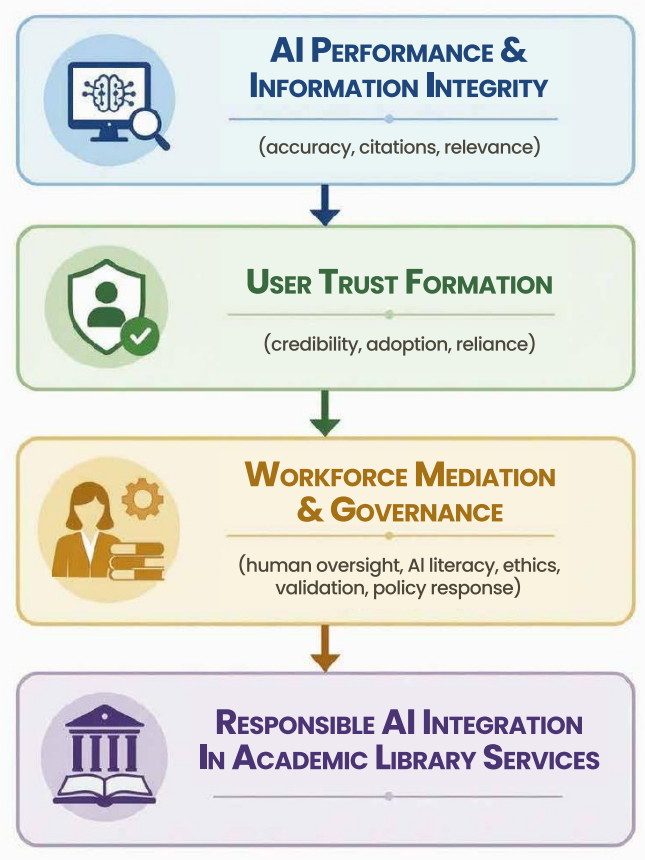


Fig. (1). Conceptual framework.

2. METHODS

2.1. Research Design

This study used a Systematic Literature Review (SLR) design framework to analyse the effects of Generative Artificial Intelligence (GenAI) on the practice of providing reference services in academic libraries and the workforce, in particular, the quality of reference services, the role of staff, and user trust. The systematic literature review provides a comprehensive understanding of the issue by synthesising available evidence (Lame, 2019). This was a review in compliance with the established PRISMA principles of evidence-based information science protocols employed to minimise selection bias and provide methodological transparency.

2.2. Search Strategy

The search strategy was aimed at ensuring that the best possible coverage of empirical evidence is given and at the same time ensuring transparency, reproducibility and relevance to library and information science. The available evidence was found in two databases, one of which was Scopus and the other was ScienceDirect. These databases were chosen to provide balance between the depth of discipline and coverage of citation and the access to both the LIS-specific and interdisciplinary studies. The selection of these databases was based on their comprehensive coverage of library and information science, information systems, and higher education research, as well as their exceptionally high indexing of peer-reviewed research journals. The searches were restricted to publications from 2023 to 2026 in the topical area, as large language models have advanced rapidly and require the inclusion of reputable publications in the search field. This temporal boundary ensures that findings reflect the post-ChatGPT phase of generative AI development. Appropriate keywords with Boolean operators were used to identify relevant evidence. Various search terms were used in both selected databases. Scopus included (ChatGPT OR 'AI chatbot*' OR 'generative AI') AND ('reference librarian*' OR 'human reference' OR 'traditional reference') AND (comparison OR comparative OR evaluation OR performance), leveraging Scopus's structured indexing and quality control for comparative empirical studies. ScienceDirect included ('generative AI' OR ChatGPT OR 'AI-assisted system') AND ('librarian' OR 'job role'). The database search results are described in Appendix **A**. To guarantee a consistent interpretation and a similar methodological appraisal, only English publications were taken into account. Duplicate records were identified and removed before the screening stage.

2.3. Study Selection Criteria

2.3.1. Inclusion Criteria

The studies that met a set of clearly specified eligibility conditions were considered in the review process. The studies had to be eligible and contain primary empirical results of qualitative, mixed-methods, or quantitative research designs and the studies that explicitly examined generative AI or large language model-based systems (ChatGPT or GenAI). The research setting also had to be within academic libraries, academic information services or closely related, highly academic settings. Studies reported at least one of the core outcomes, reference or answer quality, citation accuracy, librarian or staff roles and competencies, or user trust and credibility were included. Only studies published between 2023 and 2026 in peer-reviewed journals were included to ensure both rigour and relevance.

2.3.2. Exclusion Criteria

Review articles, systematic reviews, conceptual papers, editorials, opinion pieces, and policy commentaries were excluded. Pure technical AI studies that did not have a direct relationship with the services of an academic library or the information work in general were excluded. Research studies that did not exhibit methodological transparency, empirical evidence, or connection to library services or workforce implications were excluded. To maintain a contextually and analytically consistent research process, studies based on K-12 education (education from kindergarten to 12th grade) alone and those based on corporate knowledge management or non-academic public library environments were excluded.

2.4. Data Extraction

A data-extraction framework was created to enable comparability between the studies that were included. The information extracted included bibliographic information

(author, year, country), study design, sample characteristics, library service context, and outcome variables. Special care was given to reference quality (*e.g.,* the accuracy, completeness, citation reliability), workforce (*e.g.,* redistribution of work, skills demanded, professional identity), and user trust (*e.g.,* perceived credibility, reliance, willingness to trust AI-mediated services) indicators. Critical findings were documented in a systematic way.

2.5. Data Synthesis and Analysis

The diversity of the study designs and outcome measures made it necessary to use the narrative and thematic synthesis methodologies. The primary categorisation of studies was on the focus of analysis: quality of references, assessments on the workforce, or reliability to the users. The later identification of cross-cutting themes allowed the analysis of cross-domain interactions such as mediation of workforce and trust, or reference accuracy and professional role redesign. The quality of methodology was evaluated to determine the robustness of the evidence base and to establish long-term research gaps.

2.6. Ethical Considerations

The proposed study is limited to publicly available secondary data. Therefore, no ethical approval was necessary. The credibility of the included studies was ensured by providing correct citations, documenting all procedures, and reporting the true findings. The inclusion criteria were adhered to and analytical procedures were followed systematically throughout the review.

3. RESULTS

3.1. Data Screening

The study selection process followed the PRISMA framework to ensure methodological transparency and consistency throughout the screening stages (Fig. **2**). An initial database search identified 577 records from Scopus (n = 257) and ScienceDirect (n = 320). Before screening, 40 duplicate records were removed, leaving 537 records for title and abstract screening. During this stage, studies not specifically related to generative AI in academic library services were excluded (n = 150), alongside 12 non-English publications, 285 records comprising books, editorial papers, and conference proceedings, and 60 doctoral dissertations, and secondary reviews.

Following the screening stage, 30 reports were sought for retrieval. However, 9 reports were excluded because full texts were unavailable through open-access sources or institutional archives. Consequently, 21 full-text articles were assessed for eligibility. At the eligibility stage, 8 studies were excluded because they did not sufficiently address at least one of the core review dimensions, namely reference quality, librarian workforce roles, or user trust. Ultimately, 13 empirical studies satisfied all inclusion criteria and were included in the final systematic review (Table **2**).

3.2. Quality Assessment and Risk of Bias

Overall, the quality appraisal indicates moderate-to-high methodological quality, with recurrent limitations in recruitment transparency, representativeness, and reliance on self-report. The cross-sectional and descriptive evidence comprised (Chigwada & Pasipamire, 2024; Deschênes & McMahon, 2024; Haris *et al.,* 2025; Ismail *et al.,* 2024; Elsayed & Abusharhah, 2025; Khan, 2025; Khan *et al.,* 2026), and the survey findings summarised by (Grams, 2024). These studies were appraised with the CASP checklist (Appendix **B**). Recruitment strategies were frequently unclear or convenience-based. The survey summarised by (Grams, 2024) was limited by a small single-institution sample, while (Chigwada & Pasipamire, 2024) also had a small sample and limited analytical depth, increasing the risk of sampling bias.

The qualitative and documentary studies by (Chen, 2026; and Huang *et al.,* 2023) were appraised with the CASP Qualitative Checklist. Both had clear aims and coherent designs. Chen's interview study provided rich thematic findings but did not clearly report researcher-participant reflexivity. (Huang *et al.,* 2023) used a transparent comparative sample of strategy documents from 50 universities, although limited reporting of coding reflexivity constrained assessment of interpretive influence. (Gmiterek & Kotuła, 2025) was the sole mixed-methods study and met most MMAT criteria, with limitations relating to institutional diversity and control of confounding factors. (Lai, 2023; and Mugaanyi *et al.,* 2024) were appraised with the JBI quasi-experimental checklist. Both lacked a control group; pre/post measurement and follow-up were unclear for both; and Lai's statistical analysis was rated only partially appropriate. These limitations support a moderate-to-high, rather than uniformly high, confidence rating and require cautious interpretation of generalisability and long-term institutional impact.

Importantly, no studies were excluded based on quality appraisal criteria, as all met the lowest methodological acceptability standards established by CASP, MMAT, or JBI. In turn, quality assessment was used to give greater weight to the interpretation of findings rather than to establishing eligibility. Experimental and quasi-experimental studies were accorded greater evidentiary weight in assessing reference accuracy and bias than survey-based and qualitative studies, which were more cautious in extrapolating to workforce preparedness or user confidence when the study context was not replicated. The reference to limitations identified, especially the non-probability sampling method, representativeness, and the unavailability of longitudinal designs, helped to understand generalisability constraints. To that end, the synthesis of results focuses on convergent patterns across methods and geographical locations, rather than on single-study influences, so that the conclusion reflects the strengths and consistency of the relationships within the evidence base rather than a single high-impact report. Across the included studies, the principal risks of bias included convenience and volunteer sampling, self-reported behavioural measures, limited representativeness, and the absence of longitudinal validation. Response bias and social desirability bias were highlighted in survey-based studies, but the absence

of reflexivity reporting meant that there was moderate interpretive bias in qualitative studies. Experimental studies had similar evidence of not being subject to measurement bias, albeit specific to their contexts with limited external generalisability. No evidence of selective outcome reporting was found. Quality assessments for each study, as per their designated appraisal tool, along with an overall summary of quality assessments, are provided in Appendix **C**.

Table 2. Characteristics of 13 selected studies.

Author (Year)	Context	Method	What Was Tested	Key Quantified Findings	Core Outcome
(Chen, 2026)	Taiwan University libraries	25 librarian interviews	ChatGPT for reference services	100% required human oversight; effective only for simple queries	ChatGPT is assistive, not autonomous
(Chigwada & Pasipamire, 2024)	LIS students (Zimbabwe)	Survey (n = 59)	ChatGPT for academic work	>70% demanded AI-literacy training	Users rely on AI despite ethical & accuracy risks
(Deschênes & McMahon, 2024)	Harvard students	Survey (n = 360)	ChatGPT for research	64% use; 66% do not trust outputs	High use, low credibility
(Elsayed & Abusharhah, 2025)	Arab university libraries	Survey (n = 272)	AI in library services	37.5% use AI; only 12% met ethical issues	Low readiness, weak governance
(Gmiterek & Kotuła, 2025)	Polish university libraries	Mixed-methods	GAI in libraries	46% use GAI; 7% have policies	Adoption without regulation
(Grams, 2024)	Nigerian university students	Evidence summary of survey (underlying n = 54)	ChatGPT compared with reference librarians	98% valued speed; 87% still consulted librarians; 66.6% noted possible incorrect responses	ChatGPT supplements rather than replaces human reference support
(Haris *et al.,* 2025)	Academic library users	Survey (n = 383)	AI perception & adoption	Optimism towards AI services = 3.99/5; no significant gender or age differences	Perceived usefulness drives adoption
(Huang *et al.,* 2023)	Top 25 UK and top 25 Mainland China universities	Comparative documentary/content analysis	AI in university and academic-library strategy	AI rarely explicit in UK strategies; more visible in Chinese university visions; library strategies seldom prioritised AI	Strategic planning lagged behind AI opportunity and practice
(Ismail *et al.,* 2024)	Pakistan libraries	Survey (n = 150)	AI literacy & readiness	AI interest = 4.61/5; weak infrastructure	Motivation > capacity
(Khan, 2025)	University librarians across GCC countries	Descriptive quantitative survey (n = 160)	GenAI adoption, applications, training, and challenges	Widespread ChatGPT/Gemini/Copilot use; major training and institutional-support gaps	High adoption but uneven organisational readiness
(Khan *et al.,* 2026)	Academic library professionals in South Asia, Africa, and the Middle East	Cross-sectional survey and SEM (n = 305)	Ethical risk, competence, governance, and responsible use	Ethics competence and perceived risk strengthened governance; governance increased responsible-use intentions	Responsible adoption depends on competence, policy, and governance
(Lai, 2023)	Academic music library	Experimental	ChatGPT accuracy	Performs poorly on complex queries	Unsafe for high-level reference
(Mugaanyi *et al.,* 2024)	Academic citations	Experimental	ChatGPT citations	Only 8–33% DOIs accurate	High hallucination risk

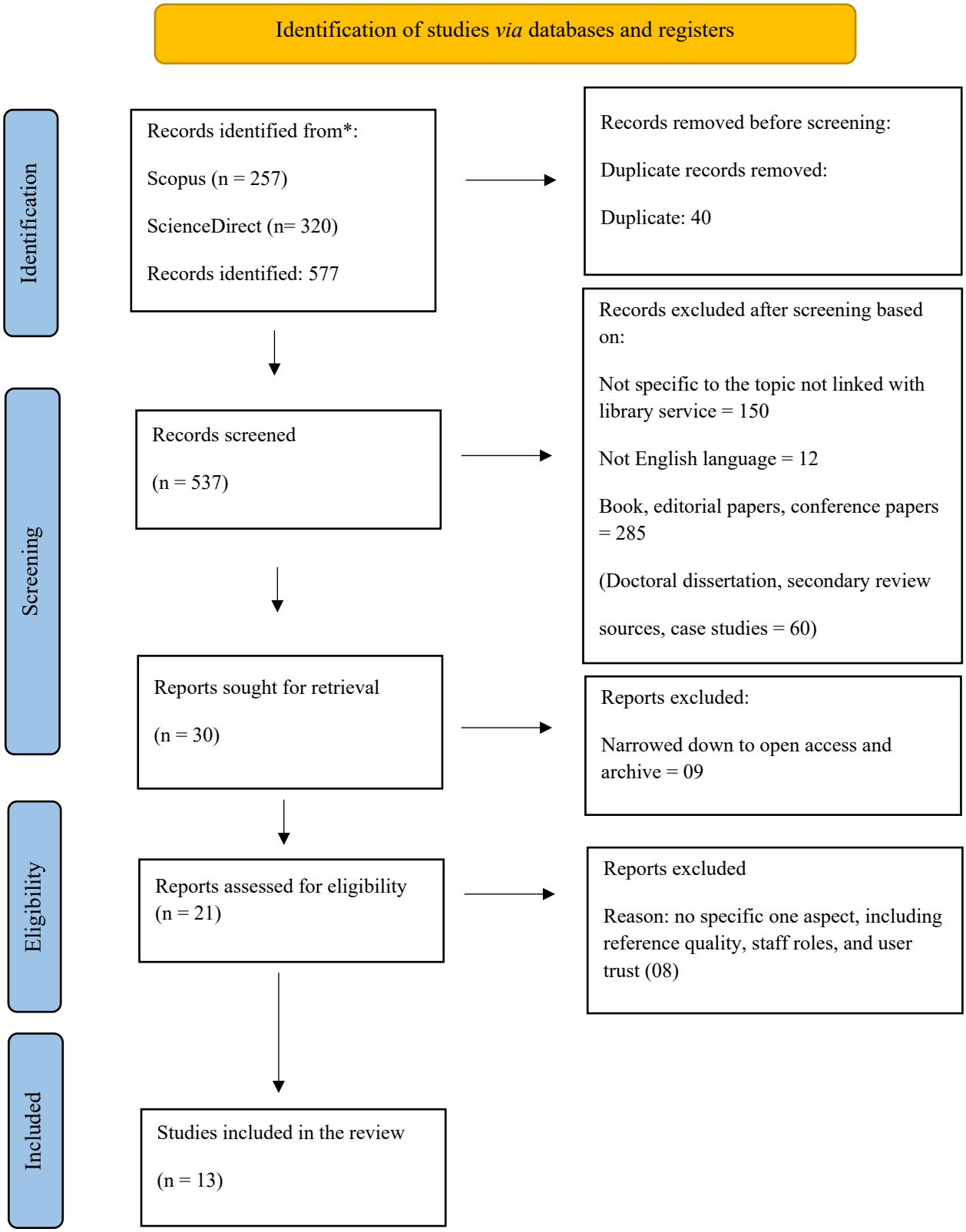


Fig. (2). PRISMA flowchart.

3.3. Study Characteristics

The articles included in this review have a diverse methodology and analyse the field of generative AI and its use in the environment of academic library service, workforce practice, and user trust. The study designs of the 13 major studies include experimental studies, large-scale surveys, mixed-methods studies, and qualitative case studies, thus demonstrating the complexity of using AI in libraries (Table **2**). The quality of references and reliability are also the subject of a significant percentage of the studies, especially through comparative indicators of the performance of ChatGPT compared to its actual or simulated reference queries.

The included studies addressed diverse dimensions of generative AI integration in academic libraries, including reference accuracy, citation reliability, workforce transformation, governance preparedness, AI literacy, user trust, and adoption behaviour. Experimental studies primarily examined information integrity and reference quality, while survey-based and qualitative studies focused on professional readiness, governance concerns, and user perceptions across different institutional and regional contexts.

3.4. Findings

The results section is explicitly structured around a socio-technical integration model in which reference service quality outcomes shape user trust and, in turn, necessitate workforce mediation and governance responses. This model provides the analytical framework through which empirical findings are organised, interpreted, and synthesised across studies, ensuring that quality, labour, and trust are examined as interdependent rather than isolated dimensions.

3.4.1. Scope and Distribution of the Evidence Base

The review is anchored on 13 peer-reviewed studies published from 2023 to 2026. Experimental or quasi-experimental studies by (Lai, 2023; and Mugaanyi *et al.,* 2024) evaluated reference performance and citation reliability. Eight survey-based or descriptive studies. (Chigwada & Pasipamire, 2024; Deschênes & McMahon, 2024; Elsayed & Abusharhah, 2025; Grams, 2024; Haris *et al.,* 2025; Ismail *et al.,* 2024; Khan, 2025; and Khan *et al.,* 2026) examined adoption, trust, ethical risk, AI literacy, and workforce readiness. (Chen, 2026; and Huang *et al.,* 2023) provided qualitative interview and documentary evidence, respectively, while (Gmiterek & Kotuła, 2025) used mixed methods. The studies span Europe, Asia, the Middle East, Africa, and the Global South. Their thematic areas include reference quality, librarian roles and competencies, strategy and governance, and user trust and adoption behaviour. Together, the evidence base shows how performance constraints, user perceptions, institutional strategy, and workforce mediation interact.

3.4.2. Impact of Generative AI on Reference Quality and Information Integrity

In both the experimental and evaluative literature, a clear performance gradient is evident in generative AI-mediated reference services, and effectiveness depends heavily on question complexity. Experimental research using controlled testing with absolute reference questions shows that ChatGPT performs relatively well on low-complexity queries (directional, facilities-related, policy, and simple factual). In an experimental assessment of 58 real academic library reference queries, (Lai, 2023), the data indicate that accuracy was highest for low-READ-scale queries (Table **3**). In contrast, in a qualitative study by (Chen, 2026), all librarians agreed that low-READ-scale queries could be effectively handled by AI, thereby reducing routine informational queries.

In comparison, high-complexity reference tasks cause a steep decline in performance. (Lai, 2023) notes that ChatGPT was rated lowest in higher research queries, known-item searches, and access to e-resources, particularly when a specific institutional context or disciplinary complexity was required. The survey evidence summarised by (Grams, 2024) is consistent with this pattern: 72.3% of students agreed that ChatGPT could not comprehend some questions and 66.6% agreed that its responses could be incorrect. The librarians interviewed in (Chen, 2026) also reported being reluctant to use AI in high-stakes reference interactions due to the risk of misinterpretation and partial contextual reasoning. The experimental methods show that retrieval accuracy decreases in proportion to the complexity of a query, and the qualitative data from practising librarians support this notion, indicating that these performance deficiencies are due to structural limitations and not temporary conditions. These quality limitations directly increase reliance on librarian mediation and shape users' scepticism towards AI-mediated reference outputs.

Alongside the superficial quality of the answer, information integrity failures pose the most significant threat to AI generation in the academic reference domain. There is also quantitative experimental evidence of high percentages of citation hallucination and bibliographic inaccuracy. In the cross-disciplinary citation experiment by (Mugaanyi *et al.,* 2024), only 32.7% of DOI references in natural sciences and 8.5% in humanities outputs were accurate, and hallucinated references in humanities-oriented outputs reached 89.4% (Table **4**). These disciplinary differences show that generative AI is weakest in areas where citation conventions are heterogeneous and context-specific. These integrity failures help explain the relationship between citation inaccuracy, user scepticism, and governance concerns, linking reference quality directly to trust erosion and governance risk.

These results build upon previous issues raised in reference-service assessments. (Lai, 2023) also identified fabricated or misleading references in responses to research-oriented questions, but (Chen, 2026) states that librarians cited hallucinations as one of the primary reasons they should not use advanced AI in reference services. Combined, these studies demonstrate that plausibility rather than verifiability governs AI outputs another epistemic problem that the standards of academic libraries cannot accept. In short, experimental studies reveal that gains in speed and accessibility are achieved at the cost of epistemic reliability, a trade-off that survey-based optimism alone fails to capture. Despite the fact that generative AI helps to optimise the process and make it more user-

friendly, it also threatens the principles of verification according to which scientific librarianship is established. (Deschênes & McMahon, 2024) disagree with the idea of unsupervised AI-based reference services, especially those managed by users who do not possess the necessary experience to identify the flaws. As a result, the necessity of the human validation layers can be found in the literature. Instead of doing away with reference services, (Chen, 2026; and Huang *et al.,* 2023) found that AI remained weakly embedded in most formal academic-library strategies.

3.4.3. Transformation of Librarian Roles and Workforce Competencies

Generative AI redeploys librarian labour not as a proactive innovation strategy, but as a compensatory response to documented weaknesses in reference quality and information integrity. According to the interview-based results provided by (Chen, 2026), 100% of the reference librarian respondents (n = 25) stressed the importance of human validation when using ChatGPT, especially for complex or high-stakes queries, which would make librarians central filters of quality rather than dispassionate intermediaries. (Khan, 2025) found that GCC university librarians were already using GenAI for research queries, article summarisation, information retrieval, research writing, and service enhancement, but many depended on self-directed training because formal programmes and institutional support were uneven. (Khan *et al.,* 2026) further showed that AI ethics competence and ethical governance significantly strengthened responsible-use intentions among 305 academic-library professionals. Similar findings were reported by (Ismail *et al.,* 2024) in Pakistani academic libraries, where respondents expressed high levels of interest in AI integration (4.61/5) despite persistent infrastructural and training limitations, suggesting that institutional readiness remains lower than professional motivation. According to (Gmiterek & Kotuła, 2025), only 46.4% of Polish academic libraries today use generative AI but only 7.1% have official guidance on AI, and 77.8% cite the lack of employee competence as the most significant challenge. While survey studies report strong professional optimism, they simultaneously expose governance and skills gaps that limit safe operationalisation, revealing a disconnect between aspiration and institutional readiness. (Elsayed & Abusharhah, 2025) found that although 81% of respondents acknowledged potential ethical risks associated with AI, only 12% reported direct institutional experience managing AI-related ethical issues, suggesting a gap between ethical awareness and operational preparedness. The signs of misalignment can be seen at an early stage in the career ladder; according to (Chigwada & Pasipamire, 2024), over 70% of LIS students mentioned repeated misinformation, bias, and ethical uncertainty despite the heavy use of ChatGPT, which indicates that early adoption moves up the career ladder faster than formal teaching on AI literacy can do. These workforce pressures emerge directly from trust-sensitive reference risks, illustrating how declining epistemic reliability necessitates increased human mediation rather than automation (Fig. **3**).

One weakness identified is the lack of formal AI governance structures. (Gmiterek & Kotuła, 2025) also note that there are few libraries that document AI policies, which is associated with disjointed, uncoordinated, and spontaneous experiments. This policy vacuum undermines the continuity of services, professional trust, and accountability, particularly when AI products are incorporated into academic scholarship.

3.4.4. User Trust, Perceived Credibility, and Adoption Behaviour

Empirical research involving end-users reveals a strong use-trust paradox rooted in uneven reference quality and limited transparency of AI outputs. (Deschênes & McMahon, 2024) state that about 64–65% of students had viewed or planned to use generative AI in academic work. Still, 65% had found AI-generated work untrustworthy enough to use in academia. The fear of information integrity was particularly high, as 88.6% of respondents reported fear of fake information, and 83.1% reported unclear or unreliable sources. Thus, cross-validation with references to trusted data sources or human oversight was a regular practice among many users. This behaviour reflects pragmatic dependency driven by convenience, rather than epistemic confidence grounded in verified accuracy. However, trust is a critical factor in the level of adoption. Trust emerged as a decisive factor influencing AI adoption. Although audit-based studies mitigate concerns about demographic bias, they do not address the dominant trust barrier identified in experimental studies: accuracy and verifiability. The diffusion-oriented evidence also provides further context in terms of the dynamic as (Haris *et al.,* 2025) in their survey of 383 users of academic libraries found a high average optimism score about the potential use of AI in enhancing library services (3.99/5) and no differences by gender or age, suggesting that the perception of usefulness, rather than demographics, drives the path of adoption (Table **5** and Fig. **4**).

High levels of audit data represented an important reassurance regarding concerns about demographic bias. For example, (Huang *et al.,* 2023) demonstrated that academic libraries in the United Kingdom and Mainland China are embedding artificial intelligence within institutional strategies, although implementation priorities differ considerably. UK libraries primarily emphasised governance, staff capability, and responsible AI adoption, whereas Chinese institutions focused more strongly on technological innovation and digital transformation. These findings indicate that successful implementation depends not only on technological capability but also on institutional planning, workforce preparedness, and governance structures.

3.4.5. Integrative Synthesis: Interdependence of Quality, Workforce, and Trust

Generative AI, in turn, can be considered a socio-technical institution where failures observed in the reference during the experimentation process will discredit the user, leading to the need to mediate and indirect the workforce to the necessary responses. Altogether, the application of generative AI is promising yet operationally constrained as a source of

academic information when used in a non-monitored mode. Transparency, accuracy, and institutional framing are sources of user trust, thus requiring that sustainable integration harmonises AI application with professional values, governance models, and evidence-based standards of service (Fig. **5**). Nevertheless, the analysed evidence also indicates the existence of several paradoxes between contexts and methodologies. The literature on surveys (*e.g.,* Haris *et al.,* 2025) shows high levels of hope for the use of AI tools and high perceived usefulness of those tools. However, the experimental studies (*e.g.,* Lai, 2023; Mugaanyi *et al.,* 2024) show significant gaps in citation reliability and contextual accuracy. Similarly, limited evidence of demographic bias was noted for LLM-driven reference services by (Huang *et al.,* 2023), while concerns about continuing professional responsibilities for ethical conduct and misinformation risks were accentuated by qualitative research reports from (Chen, 2026). These differences imply that user optimism might be due to perceived convenience, not to actual epistemic confidence.

However, there is also contextual variation within regions and their institutional contexts. While European and North American studies tended to concentrate on aspects of governance maturity and the integration of AI, studies conducted in Pakistan and Arab institutions and in the Global South were more focused on infrastructural constraints, policy gaps, and lack of preparedness for AI literacy. In addition, gaps in citations were more evident in the humanities areas than in the science areas, demonstrating differences related to the type of domain where AI might be used when providing a reference.

Table 3. The average quality of chatGPT's answers based on question complexity using the read scale. Source: (Lai, 2023).

Question Complexity	Quality			Overall Average Quality
	Completeness	Accuracy	Further Assistance	
READ level 1 (n = 3)	3.00	2.67	3.00	2.89
READ level 2 (n = 25)	2.44	1.76	1.92	2.04
READ level 3 (n = 16)	2.50	1.81	1.69	2.00
READ level 4 (n = 12)	2.50	1.67	1.92	2.03
READ level 5 (n = 2)	3.00	1.50	2.00	2.17
Overall	2.52	1.79	1.91	2.07

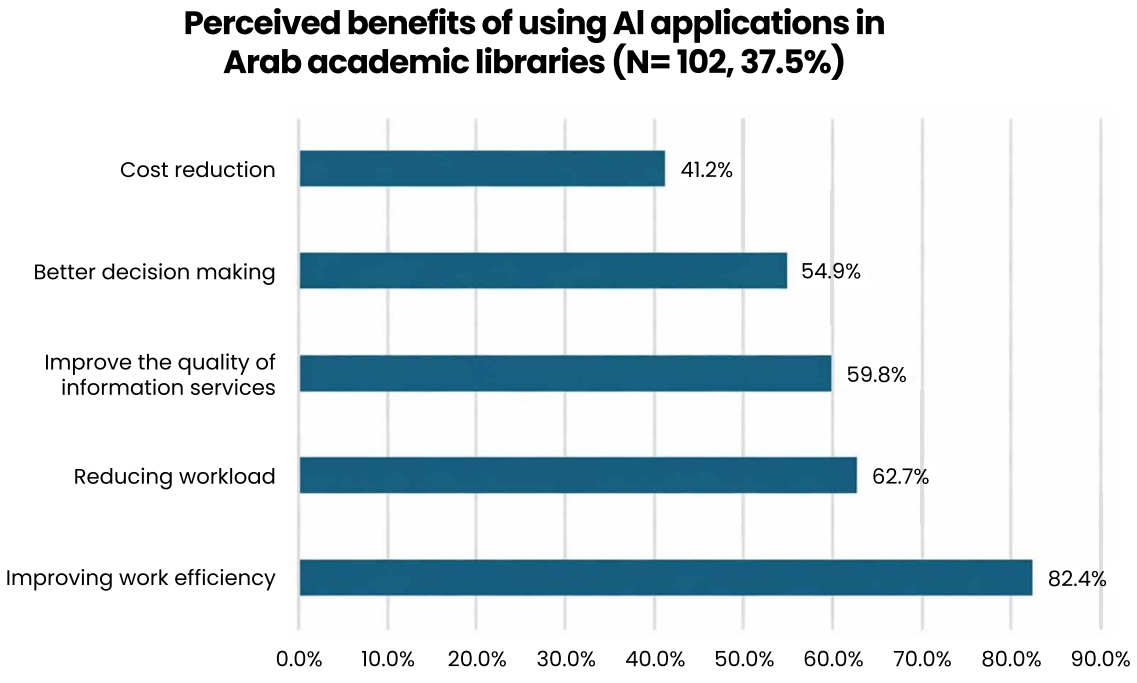


Fig. (3). Perceived benefits of using AI applications in Arab academic libraries. Source: (Elsayed & Abusharhah, 2025).

Table 4. Data analysis results. Source: (Mugaanyi *et al.*, 2024).

Variables	Natural Sciences (n = 55)	Humanities (n = 47)	*P* value[a]
Citation exists, n (%)	40 (72.7)	36 (76.6)	.42
Citation accurate, n (%)	37 (67.3)	29 (61.7)	.35
Relevant, n (%)	39 (70.9)	35 (74.5)	.43
DOI[b] exists, n (%)	39 (70.9)	18 (38.3)	.001
DOI accurate, n (%)	18 (32.7)	4 (8.5)	.003
DOI hallucination, n (%)	34 (61.8)	42 (89.4)	.001
Levenshtein distance, mean (SD)	64.13 (42.26)	42.15 (40.23)	.009

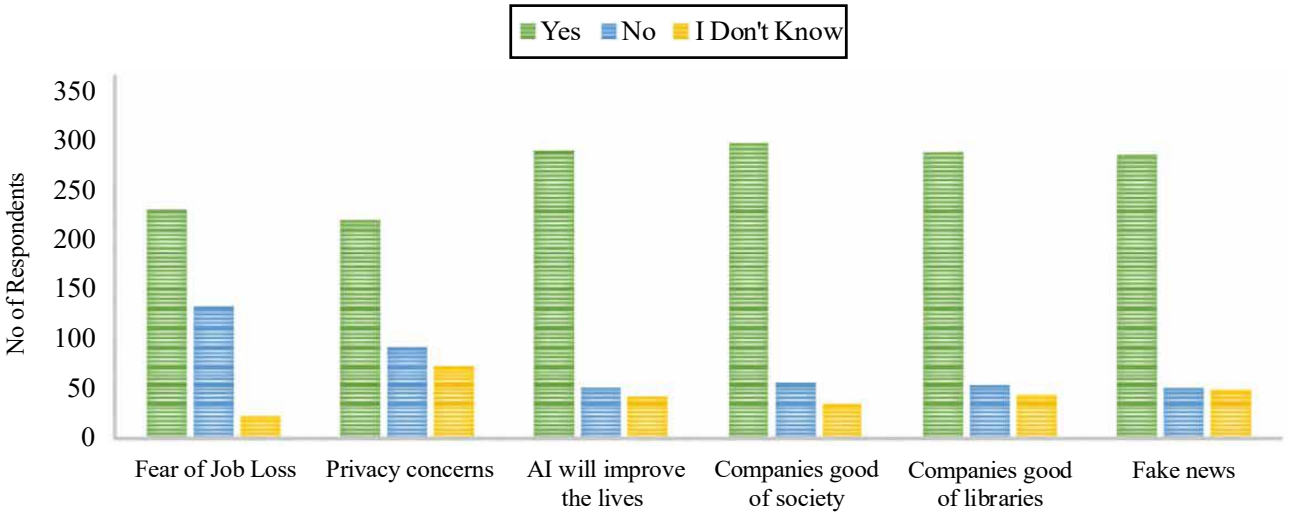


Fig. (4). Respondents' concern regarding Artificial Intelligence. Source: (Haris *et al.,* 2025).

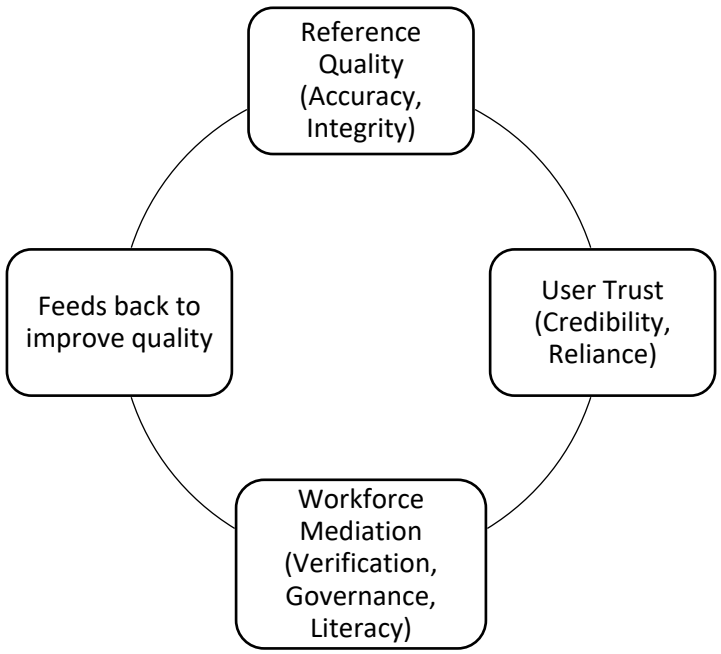


Fig. (5). Socio-technical integration model linking reference quality, workforce mediation, and user trust in AI-enabled academic libraries.

Table 5. Respondents' knowledge and optimism about Artificial Intelligence. Source: (Haris *et al.*, 2025).

Respondents' Knowledge and optimism about AI		
Topic	**Mean**	**SD**
Conceptual knowledge of AI	3.44	1.45
Knowledge of trends in AI	3.38	1.45
An optimistic perspective on the potential of AI to enhance library services	3.99	1.56

4. DISCUSSION

This systematic review provides a synthesised and critical insight into the transformation of academic library services, professional functions, and user confidence brought about by generative AI, specifically large language models like ChatGPT. Rather than demonstrating complete technological replacement, (Lai, 2023; Grams, 2024; and Mugaanyi *et al.,* 2024) suggest that generative AI restructures the epistemic organisation of academic library services by shifting professional emphasis from information retrieval towards verification, mediation, and governance. It depends on the task's complexity, the organisation, and the role of humans, which reiterates broader socio-technical views in information science. These results align with the findings of (Bawden & Robinson, 2012), who identified that automated tools are better at performing surface-level retrieval but struggle with the interpretive and evaluative aspects of information work. These findings reinforce broader concerns regarding epistemic reliability in probabilistic language systems. (Ji *et al.,* 2023) findings also align with this review; they identified that hallucination is a natural property of probabilistic language models and is not introduced by implementation failure. Collectively, these findings suggest that scaling AI systems alone does not resolve epistemic reliability concerns, thereby challenging techno-optimist assumptions regarding autonomous academic information services. Instead, they support the argument raised by (Floridi *et al.,* 2018), to the effect that epistemic reliability in information systems relies on human-in-the-loop validation, particularly where authority, accountability, and scholarly credibility are demanded. The implementation of AI in academic libraries, where reference services often serve as epistemic gatekeeping systems, risks undermining the fundamental professional and institutional values.

In addition, the evidence implies that professional labour has to be reorganised, and the role of librarians is increasingly repositioned as AI mediators, validators, and ethical stewards rather than being replaced. This is consistent with the theory of skilled jurisdiction as developed by (Abbott, 1988) that is the response of professions to the change in technologies to renegotiate tasks but not to delegate power. The transition of transactional reference work to oversight, AI literacy education, and governance is in line with the trends outlined by (Cox *et al.,* 2019). These authors also observe that there is a need to study other aspects of the digital transformation like the development of discovery systems and algorithmic recommender systems. However, (Gmiterek & Kotuła, 2025; and Elsayed & Abusharhah, 2025) indicate a chronic capacity and governance gap. Despite the high interest and perceived benefits (Fig. **3**), most institutions lack formal AI policies, and staff skills are uneven. This aligns with other criticisms of the effects of policy lag in educational technologies, as noted by (Rafiq-uz-Zaman, 2025), who argues that a new resource is adopted without the creation of regulatory and ethical protections. The lack of well-developed institutional frameworks places an excessive burden on individual librarians for risk mitigation, accompanied by related issues of sustainability, accountability, and professional burnout. Trust appears to be the most prominent moderating factor of AI adoption in other user-friendly studies. The paradoxical nature of using both high and low at the same time, and having high and low trust, is defined by (Wen, 2024) as conditional trust; the users are pragmatic about systems but not confident in their efficacy in epistemic terms. This finding is consistent with (Fügener *et al.,* 2022), who indicate that educational and healthcare stakeholders adopt algorithmic tools to optimise efficiency but are not yet willing to do away with cognitive tasks entirely. The concept of algorithmic bias is also relevant to the review despite long-standing concerns about accuracy and verification. (Weidinger *et al.,* 2021) find limited demographic bias in contemporary large language models in the setting of academic references. These results suggest that fairness alone does not establish trustworthy relationships; an effective system of credibility in academic libraries requires transparency and traceability of the sources and compliance with academic standards.

4.1. Theoretical Contribution: From Tool Evaluation to Socio-Technical Governance

This review supports the academic discourse by providing a socio-technical account of generative AI in libraries by moving beyond tool-focused or adoption-oriented discourse. The view on generative AI in academic libraries as an innovation that saves time and resources or as a technology that needs to be ethically regulated, largely depends on current reviews, including works by (Cox *et al.,* 2019). Although important, such accounts often treat the quality of references, workforce change, and user trust as parallel issues rather than interdependent dynamics. Combining experimental, survey, and qualitative evidence, this review shows that the problem of failures in reference quality, specifically citation hallucination and contextual misinterpretation, is the central event that appears to influence erosion of trust among two-thirds of the

workforce, which in turn prompts the need to have more mediation in the workforce and institutional governance. Such a holistic perspective suggests that AI-related inaccuracies should be viewed not merely as technical shortcomings but as organisational risks requiring policy development, workflow redesign, and professional oversight. Notably, the review clarifies that librarians are repositioned as epistemic arbitrators who are required to justify knowledge assertions, critically assess AI outcomes, and protect academic values. This stance is in tandem with socio-technical explanations of information infrastructure, based on the explanation by (Orlikowski, 2007) that human actors stabilise trust in complex systems by providing supervision and norm-setting. By doing so, the review establishes a theoretical one-way link between evidence of AI performance and professional governance, providing a framework through which previously fragmented reviews could not be empirically examined.

4.2. Strengths and Limitations

The primary strength of the reviewed literature is the variety of methodologies used, as there are studies that apply experimental design (Lai, 2023; Mugaanyi *et al.,* 2024), large-scale quantitative surveys (Lai, 2023; Mugaanyi *et al.,* 2024), qualitative interviews (Chen, 2026). This methodological diversity makes the evidence base more robust, as it combines objective performance indicators with the perceptions of professionals and users and makes the interpretation of the effects of generative AI nuanced. External validity beyond Western libraries is enhanced by the broad geographic coverage of the study, which comprises North America, Europe, Asia, the Middle East, and the Global South (Chen, 2026, Ismail *et al.,* 2024, Elsayed & Abusharhah, 2025).

However, there remain several limitations. Most sources focus on ChatGPT or similar large language models (Lai, 2023; Chen, 2026; Mugaanyi *et al.,* 2024), thereby overlooking alternative or locally developed AI-based solutions. Survey-based research generally relies on self-reported perceptions, and the approach cannot be seen as precise regarding the assessment of actual behaviour or the level of competency (Deschênes & McMahon, 2024; Haris *et al.,* 2025). Typically, experimental studies are cross-sectional studies, which restrict our understanding of the longitudinal process of adaptation and learning (Lai, 2023). Moreover, although governance and skills gaps are often reported (Gmiterek & Kotuła, 2025; Elsayed & Abusharhah, 2025), empirical research has yet to test the effectiveness of individual policy or training interventions. Last but not least, although new fairness audits are emerging, the intersectional and discipline-specific trust relationships are under-researched, providing evidence for future research directions.

CONCLUSION

This systematic review explored how generative AI has affected reference services in academic libraries, workforce roles, and user trust, based on recent empirical evidence across geographic and institutional settings. The findings suggest that generative AI offers operational advantages in low-complexity reference environments but continues to face substantial limitations in citation reliability, contextual reasoning, and epistemic verification. Instead of replacing librarians, generative AI is transforming professional practices by intensifying the need for mediation, verification, ethical oversight, and AI literacy, which only strengthens the continued centrality of human expertise in the provision of academic information services. User acceptance is widespread but conditional, and trust turns out to be rather a weak but conclusive factor influenced by transparency, the credibility of the source, and the framing by institutions. Future research should focus on longitudinal studies to evaluate the long-term effects of AI integration on reference service quality, professional identity, and the nature of user-AI interactions. It is also necessary to compare the assessments of different generative AI services, such as ChatGPT, to identify which of the limitations are systemic or model-specific. In addition, empirical research on governance models, training programmes, and AI literacy education could provide practical guidance on the responsible and sustainable adoption of AI in libraries.

LIST OF ABBREVIATIONS

CASP	=	Critical Appraisal Skills Programme
GenAI	=	Generative Artificial Intelligence
HEPI	=	Higher Education Policy Institute
LLMs	=	Large Language Models
SLR	=	Systematic Literature Review

AUTHOR'S CONTRIBUTION

A.S. conceived and designed the study, acquired and preprocessed the dataset, conducted the statistical analyses and machine learning experiments, interpreted the findings, developed the proposed framework, drafted and critically revised the manuscript, and approved the final version for publication.

ETHICAL APPROVAL & INFORMED CONSENT

This study did not involve human participants, animal subjects, or any procedures requiring ethical approval.

REPORTING GUILDELINES

PRISMA guidelines have been followed in this study.

AVAILABILITY OF DATA AND MATERIALS

No new data were generated in this study.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The author declares that there are no competing interests or conflicts of interest relevant to the content of this work.

ACKNOWLEDGEMENTS

Declared none.

DECLARATION OF AI

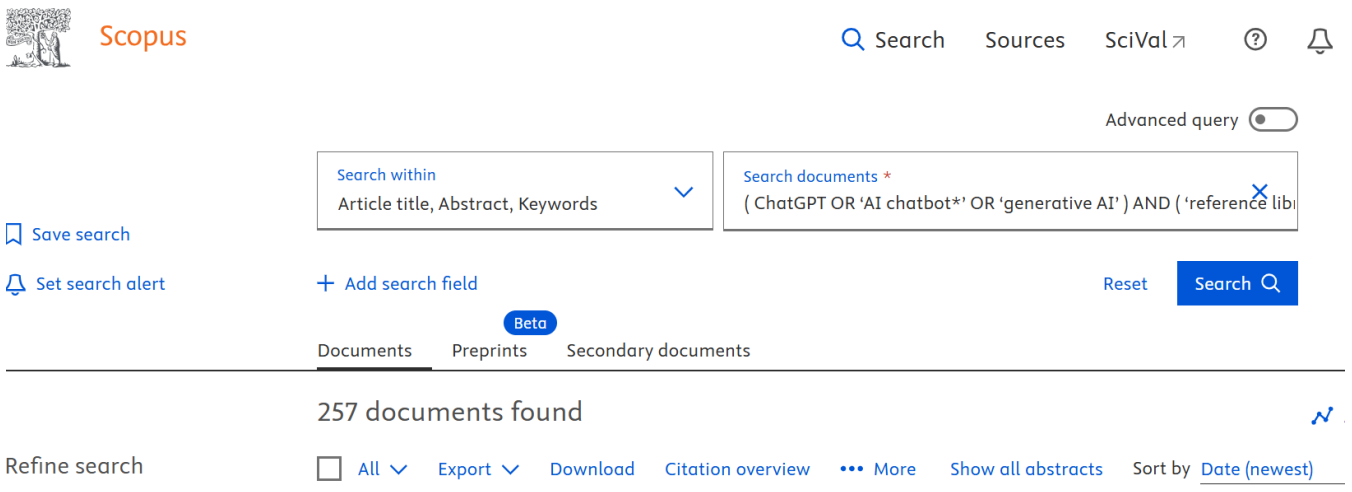
The author used ChatGPT for language editing purposes during the preparation of this manuscript. All AI-assisted content was reviewed, verified, and approved by the author, who assumes full responsibility for the final version of the manuscript.

APPENDIX

Appendix A: Database Search

Science Direct

Scopus



Search Strategy and study selection outcomes.

Database	Search Period	Records Identified (n)
Scopus	2023–2026	257
ScienceDirect	2023–2026	320
Total records identified		577

Full-text retrieval outcomes and exclusion reasons.

Retrieval Stage	ScienceDirect (n)	Scopus (n)	Total (n)	Justification
Records identified from core databases	320	257	577	Initial database search conducted across selected databases
Duplicate records removed	22	18	40	Duplicate citations identified and removed prior to screening
Records screened (title and abstract)	298	239	537	Records remaining after duplicate removal and subjected to title and abstract screening
Records excluded after title and abstract screening	280	227	507	Excluded due to lack of topical relevance (n = 150), non-English publications (n = 12), books/editorials/conference papers (n = 285), and theses, secondary reviews (n = 60)
Reports sought for full-text retrieval	18	12	30	Studies demonstrating empirical relevance and alignment with review objectives
Reports excluded due to access limitations	5	4	9	Full-text articles unavailable through open-access sources or institutional archives
Reports assessed for eligibility (full text)	13	8	21	Full-text studies evaluated against inclusion and exclusion criteria
Reports excluded after full-text assessment	5	3	8	Excluded because they did not sufficiently address reference quality, workforce roles, or user trust outcomes
Studies included in final systematic review	6	7	13	Studies meeting all inclusion criteria and contributing directly to the review synthesis

Appendix B: Quality Assessment Checklists

CASP Checklist - Cross-sectional / Descriptive studies.

CASP Criteria	(Grams, 2024)	(Deschênes & McMahon, 2024)	(Haris *et al.*, 2025)	(Ismail *et al.*, 2024)	(Elsayed & Abusharhah, 2025)	(Chigwada & Pasipamire, 2024)
1. Clear research aim	Yes	Yes	Yes	Yes	Yes	Yes
2. Appropriate methodological choice	Yes	Yes	Yes	Yes	Yes	Yes
3. Acceptable recruitment strategy	No	Can’t tell	Can’t tell	Can’t tell	Can’t tell	Can’t tell
4. Accurate measurement & validated tools	Partial	Yes	Yes	Yes	Yes	Yes
5. Data collection addressed research issue	Yes	Yes	Yes	Yes	Yes	Yes

6. Adequate sample size to minimise chance	No	Yes	Yes	Yes	Yes	No
7. Clarity of results & main findings	Yes	Yes	Yes	Yes	Yes	Yes
8. Rigorous data analysis	Partial	Yes	Yes	Partial	Yes	Partial
9. Clear statement of findings	Yes	Yes	Yes	Yes	Yes	Yes
10. Applicability to the target population	Limited	Yes	Yes	Moderate	Yes	Limited
11. Value of the research	Yes	Yes	Yes	Yes	Yes	Yes

CASP checklist — Cross-sectional / Descriptive studies (continued).

CASP Criteria	(Khan, 2025)	(Khan *et al.*, 2026)
1. Clear research aim	Yes	Yes
2. Appropriate methodological choice	Yes	Yes
3. Acceptable recruitment strategy	Can't tell	Can't tell
4. Accurate measurement & validated tools	Partial	Partial
5. Data collection addressed research issue	Yes	Yes
6. Adequate sample size to minimise chance	Yes	Yes
7. Clarity of results & main findings	Yes	Yes
8. Rigorous data analysis	Yes	Yes
9. Clear statement of findings	Yes	Yes
10. Applicability to the target population	Moderate	Moderate
11. Value of the research	Yes	Yes

CASP qualitative checklist.

CASP Criterion	(Chen, 2026)	(Huang *et al.*, 2023)
1. Clear statement of aims?	Yes	Yes
2. Qualitative methodology appropriate?	Yes	Yes
3. Research design appropriate to address aims?	Yes	Yes
4. Recruitment strategy appropriate?	Yes	Yes (top-25 universities in each region)
5. Data collected in a way that addressed the research issue?	Yes	Yes
6. Researcher–participant relationship considered?	Can’t tell	Not applicable (documentary study)
7. Ethical issues considered?	Yes	Not applicable (public strategy documents)
8. Data analysis sufficiently rigorous?	Yes	Yes
9. Clear statement of findings?	Yes	Yes
10. Value of the research?	Yes	Yes

Mixed Methods Appraisal Tool (MMAT), Version 2018.

MMAT Criteria (2018)	(Gmiterek & Kotuła, 2025)
Screening Questions (All Studies)	-
S1. Are there clear research questions?	Yes
S2. Do the collected data allow the research questions to be addressed?	Yes
1. Qualitative Component	-
1.1 Is the qualitative approach appropriate to answer the research question?	Yes (exploration of librarian perceptions and governance practices)
1.2 Are qualitative data collection methods adequate?	Yes (open-ended institutional responses and qualitative interpretation)
1.3 Are findings adequately derived from the qualitative data?	Yes (findings linked to reported institutional practices)
1.4 Is interpretation sufficiently substantiated by qualitative data?	Yes
1.5 Is there coherence between qualitative data sources, collection, analysis, and interpretation?	Yes
3. Quantitative Non-Randomised Component	-
3.1 Are participants representative of the target population?	Moderate (national academic library sample but limited institutional diversity)
3.2 Are measurements appropriate for the outcomes studied?	Yes (survey indicators measuring AI adoption and governance readiness)
3.3 Are there complete outcome data?	Yes
3.4 Are confounders accounted for in design or analysis?	Partial (limited control for institutional size and digital capacity differences)
3.5 Was the intervention/exposure implemented as intended?	Yes (consistent institutional survey implementation)
5. Mixed Methods Component	-
5.1 Is there an adequate rationale for using mixed methods?	Yes (to evaluate both adoption rates and governance perceptions)
5.2 Are qualitative and quantitative components effectively integrated?	Yes (integration of survey trends with institutional commentary)
5.3 Are outputs of integration adequately interpreted?	Yes
5.4 Are divergences between qualitative and quantitative results addressed?	Yes (contrast between AI optimism and policy limitations discussed)
5.5 Do all components meet the quality criteria of their respective traditions?	Yes (with acknowledged governance and sampling limitations)

Quasi experimental design JBI checklist.

JBI Criterion	(Lai, 2023)	(Mugaanyi *et al.*, 2024)
1. Is it clear in the study what is the 'cause' and what is the 'effect'?	✓	✓
2. Were the participants included in any comparisons similar?	✓	✓
3. Were the participants included in comparisons receiving similar treatment/care other than the exposure of interest?	✓	✓
4. Was there a control group?	✕	✕
5. Were there multiple measurements of the outcome both pre and post the intervention/exposure?	Unclear	Unclear
6. Was follow-up complete and, if not, were differences between groups adequately described and analysed?	Unclear	Unclear
7. Were the outcomes of participants included in any comparisons measured in the same way?	✓	✓
8. Were outcomes measured in a reliable way?	✓	✓
9. Was appropriate statistical analysis used?	Partial	✓

Appendix C: Summary of Quality Assessment Checklists

Summary quality appraisal.

Study Design	Appraisal Tool	Number of Studies	Overall Quality Judgment	Key Strengths Identified	Recurrent Limitations
Cross-sectional / Descriptive	CASP	8 (Chigwada & Pasipamire, 2024; Deschênes & McMahon, 2024; Elsayed & Abusharhah, 2025; Grams, 2024; Haris *et al.,* 2025; Ismail *et al.,* 2024; Khan, 2025; Khan *et al.,* 2026)	Moderate–High	Clear aims, suitable survey designs, transparent results, and appropriate descriptive or multivariate analysis	Convenience or unclear recruitment, self-report, small or single-institution samples, and partial instrument reporting
Qualitative	CASP Qualitative	2 (Chen, 2026; and Huang *et al.,* 2023)	Moderate–High	Coherent qualitative or documentary designs, clear aims, and useful thematic or comparative findings	Limited reflexivity reporting and incomplete detail on coding influence
Mixed-methods	MMAT	1 (Gmiterek & Kotuła, 2025)	Moderate–High	Integration of survey trends with qualitative institutional commentary	Limited institutional diversity and partial control of confounding factors
Quasi-experimental	JBI	2 (Lai, 2023; and Mugaanyi *et al.,* 2024)	Moderate–High	Clear exposure-outcome logic, consistent measurement, and direct assessment of AI outputs	No control groups; unclear pre/post measurement and follow-up; partially appropriate statistical analysis for Lai

REFERENCES

Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226189666.001.0001

Bawden, D., & Robinson, L. (2012). *Introduction to information science*. Facet Publishing. Available from: https://openaccess.city.ac.uk/id/eprint/3224/

Chen, S. C. (2026). Transforming Reference Services through ChatGPT: Insights from University Libraries in Taiwan. *New Review of Academic Librarianship*, *32*(1), 71–91. https://doi.org/10.1080/13614533.2025.2586271

Chigwada, J., & Pasipamire, N. (2024). Perception and use of large language models by library and information science students. *International Journal of Librarianship, 9*(3), 75–89. https://doi.org/10.23974/IJOL.2024.VOL9.3.385

Cox, A. M., Pinfield, S., & Rutter, S. (2019). The intelligent library: Thought leaders' views on the likely impact of Artificial Intelligence on academic libraries. *Library Hi Tech, 37*(3), 418–435. https://doi.org/10.1108/LHT-08-2018-0105

Deschênes, A., & McMahon, M. (2024). A survey on student use of generative AI chatbots for academic research. *Evidence Based Library and Information Practice, 19*(1), 2–22. https://doi.org/10.18438/EBLIP30512

Elsayed, A. M., & Mohammed Abusharhah, M. (2025). Artificial Intelligence adoption, perceptions, and ethical literacy among Arab academic librarians: A survey. *The Journal of Academic Librarianship, 51*, 103083. https://doi.org/10.1016/j.acalib.2025.103083

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., *et al.* (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive challenges in human–Artificial Intelligence collaboration: Investigating the path toward productive delegation. *Information Systems Research, 33*(2), 678–696. https://doi.org/10.1287/isre.2021.1079

Gmiterek, G., & Kotuła, S. D. (2025). Generative Artificial Intelligence in the activities of academic libraries of public universities in Poland. *The Journal of Academic Librarianship, 51*(3), 103043. https://doi.org/10.1016/j.acalib.2025.103043

Grams, K. (2024). Students' perspective of the advantages and disadvantages of ChatGPT compared to reference librarians. *Evidence Based Library and Information Practice, 19*(2), 130–132. https://doi.org/10.18438/EBLIP30518

Haris, M., Ansari, A. J., Malik, B. A., Lund, B. D., & Ali, N. (2025). Artificial Intelligence in academic libraries: A survey of users' perception and adoption. *Global Knowledge, Memory and Communication*, Advance online publication. https://doi.org/10.1108/GKMC-09-2024-0585

Higher Education Policy Institute. (2025). *Student generative AI survey 2025*. Available from: https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/ (Accessed on: 22 May 2026).

Huang, Y., Cox, A. M., & Cox, J. (2023). Artificial intelligence in academic library strategy in the United Kingdom and the Mainland of China. *The Journal of Academic Librarianship*, *49*(6), 102772. https://doi.org/10.1016/j.acalib.2023.102772

Ismail, M., Hussain, A., & Haseeb, A. (2024). Examining Artificial Intelligence (AI) literacy among university library professionals in Pakistan: The case of Khyber Pakhtunkhwa. *Journal of Information Management and Library Studies*, *7*(1), 113-139. Available from: https://jimls.kkkuk.edu.pk/index.php/jimls/article/view/154

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., *et al*. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Article 248. https://doi.org/10.1145/3571730

Khan, Z. I. (2025). Exploring the utilization of generative AI by librarians in higher education across the Gulf Cooperation Council (GCC) countries: Trends in adoption, innovative applications, and emerging challenges. *Journal of Librarianship and Information Science.* Advance online publication. https://doi.org/10.1177/09610006251372630

Khan, Z. I., Oladokun, B. D., Upadhyay, A. K., & Kalbande, D. (2026). Ethical concerns of generative AI in academic libraries: A cross-regional quantitative study of librarians' risk perceptions, trust, and governance practices. *Journal of Librarianship and Information Science.* Advance online publication. https://doi.org/10.1177/09610006261461309

Lai, K. (2023). How well does ChatGPT handle reference inquiries? An analysis based on question types and question complexities. *College & Research Libraries, 84*(6), 974–995. https://doi.org/10.5860/crl.84.6.974

Lame, G. (2019). Systematic literature reviews: An introduction. *Proceedings of the Design Society: International Conference on Engineering Design, 1*(1), 1633–1642. https://doi.org/10.1017/dsi.2019.169

Mugaanyi, J., Cai, L., Cheng, S., Lu, C., & Huang, J. (2024). Evaluation of large language model performance and reliability for citations and references in scholarly writing: Cross-disciplinary study. *Journal of Medical Internet Research*, *26*, e52935. https://doi.org/10.2196/52935

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies, 28*(9), 1435–1448. https://doi.org/10.1177/0170840607081138

Rafiq-uz-Zaman, M. (2025). Between adoption and ambiguity: Navigating the AI policy vacuum in Pakistani higher education. *Research Journal for Social Affairs*, *3*(6), 877-885. https://doi.org/10.71317/RJSA.003.06.0523

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., ... & Gabriel, I. (2021). Ethical and social risks of harm from language models. *arXiv* preprint arXiv:2112.04359. https://arxiv.org/abs/2112.04359

Wen, J. (2024). On Epistemic Trust. (Doctoral dissertation, Macquarie University). https://doi.org/10.25949/26020540