# AI-Enabled Information Retrieval in Libraries: Integration, Challenges, and Prospects

Hongzhuan Guo

The Library of Anhui Polytechnic University, Wuhu, 241000, China

E-mail: 44063598@qq.com

**Abstract.** With the rapid advancement of Generative Artificial Intelligence (AIGC) technology, large language models (LLMs) such as ChatGPT and DeepSeek are profoundly reshaping the ecosystem of information retrieval services in libraries. Drawing on multiple empirical studies and technical analysis literature, this paper systematically reviews the current application scenarios, technical pathways, and enabling mechanisms of AI in library information retrieval—focusing on typical practices like conversational AI-based reference consultation and personalized recommendation. It further analyzes core challenges across four dimensions: data security and privacy protection, algorithmic transparency and bias, technology adaptation and talent shortage, as well as user trust and ethical risks. Correspondingly, targeted development strategies are proposed, including the integration of the "Generative AI + Traditional Retrieval" hybrid model, strengthening data governance and ethical norm construction, improving librarians' AI literacy and user information education, and exploring the application of federated learning and edge computing. These strategies aim to advance the intelligent and inclusive transformation of libraries, and build a credible human-AI collaborative service paradigm that balances technological empowerment with humanistic care.

**Keywords:** Artificial Intelligence (AI); Library; Information Retrieval; Generative AI; Knowledge Acquisition; Intelligent Services

## 1 Introduction

Information retrieval has long been a core function of library services. Traditional retrieval systems rely primarily on keyword matching and Boolean logic—approaches that, while mature and highly stable, exhibit notable limitations in semantic understanding, multimodal processing, and personalized service delivery. Users typically require specialized search skills and familiarity with logical operators to effectively retrieve the information they need; even so, the diversity of natural language and complexity of semantic meanings mean that literal-based matching often fails to accurately capture user intent. This results in low-relevance search outcomes, excessive redundant information, and a suboptimal user experience.



A. Y. M. A. Islam et al. (eds.), *Proceedings of the 2025 International Conference on Educational Technology and Management Information Systems (ETMIS 2025)*, Advances in Computer Science Research 129,
https://doi.org/10.2991/978-94-6239-630-2_21

In recent years, generative artificial intelligence (AI) technologies exemplified by ChatGPT and DeepSeek have advanced by leaps and bounds. Their robust natural language processing (NLP) and content generation capabilities have driven revolutionary changes in library information retrieval [1]. These large language models (LLMs) can deeply interpret the semantic context of user queries, transcending rigid literal keyword matching to substantially enhance retrieval accuracy and efficiency. For instance, users may pose ambiguous questions, engage in multi-turn conversational interactions, or provide corrective feedback in natural language; the system can then dynamically adjust its retrieval strategies to deliver responses that align more closely with user intent.

Furthermore, generative AI has unlocked the potential of cross-modal retrieval. Modern libraries now curate increasingly diverse collections, encompassing not only textual resources but also images, audio, video, and other multimedia formats. AI technologies enable the unified representation and retrieval of information across these different modalities—such as retrieving relevant images via textual descriptions or locating textual materials through visual content—effectively building a new knowledge service ecosystem characterized by "universal retrievability" of all resources.

In the realm of personalized services, AI systems can proactively deliver precise content recommendations and knowledge guidance by analyzing users' historical behaviors, preference patterns, and contextual needs. This shifts the service paradigm from "users searching for information" to "information finding users," which not only improves the efficiency of users' information access but also significantly elevates the intelligence level of library knowledge services and overall user satisfaction.

In summary, generative AI technologies are progressively reshaping the entire ecosystem of library information retrieval. Beyond driving remarkable advances in efficiency and accuracy, these technologies have catalyzed fundamental transformations in retrieval dimensions, interaction methods, and service models—injecting new vitality into traditional knowledge service institutions and defining new directions for their development.

## 2 Current Status of AI Applications in Library Information Retrieval

### 2.1 Retrieval Empowerment by Generative AI

Generative AI, leveraging its advanced semantic understanding and contextual reasoning capabilities, has driven a paradigm shift in information access—from traditional keyword-based retrieval to an intelligent question-answering mode [2]. Unlike conventional search engines that return lists of potentially relevant links or document snippets based on keyword matching, Generative AI can directly parse natural language user queries, perform context-aware logical inferences, and generate well-structured, coherent responses. This shift eliminates the need for users to manually sift through and synthesize information from massive result sets; instead, they can

obtain clear, targeted answers directly, significantly enhancing both the efficiency and quality of information access experiences.

In this evolution, generative AI models such as ChatGPT and GPT-4 have played a pivotal role. Built on large-scale pre-trained language models, these tools excel at capturing complex semantic relationships—enabling them to recognize not only the literal meaning of a query but also its implicit intent, while maintaining semantic coherence and logical consistency in generated outputs. They perform exceptionally well in unstructured tasks such as emotional support, career planning, and personal development [3], acting as listeners, advisors, or mentors to deliver empathetic and practical responses—making them highly valuable in scenarios like psychological counseling, career guidance, and creative inspiration.

For example, specific Case: Psychological Counseling Scenario: Woebot and ChatGPT Assist in Emotional Regulation Woebot, a mental health chatbot, leverages generative AI technology combined with Cognitive Behavioral Therapy (CBT) to provide users with daily emotional support. For instance, a college student suffering from test anxiety confided in Woebot: "I always feel that no matter how hard I try, I can't do well in exams, and I'm under a lot of pressure." The system not only identifies emotional signals of "anxiety" and "self-negation" but also builds trust through empathic responses (e.g., "It sounds like you've put in a lot of effort yet feel unrecognized, which must be really tough"), and guides users to record negative thoughts and restructure their cognition. Similarly, some psychological counselors use GPT-4 to assist in drafting intervention plans and quickly generate personalized conversation scripts, thereby improving service efficiency.

However, generative AI is far from omnipotent: it faces notable limitations in highly specialized, rigor-intensive fields including academic research, legal consulting, and medical diagnosis. Rooted in a probabilistic generation mechanism, these models may produce "hallucinations"—plausible-sounding outputs that contain factual inaccuracies or entirely unsubstantiated claims. For this reason, in academic writing, scientific research, and evidence-based tasks, generative AI is typically paired with traditional retrieval systems (e.g., academic databases, professional knowledge graphs, and literature retrieval platforms) to form complementary workflows. Researchers can first use generative AI to rapidly organize research ideas, gather background information, or draft preliminary literature abstracts, then verify information sources and trace original literature via professional retrieval systems to ensure the accuracy and credibility of their work.

It is evident that generative AI has not replaced traditional retrieval tools outright; instead, it has expanded the boundaries of human-computer interaction and reshaped the landscape of knowledge services. It enhances information access efficiency for general users while collaborating with traditional technologies in specialized domains to build more robust and reliable human-AI collaboration mechanisms. Looking ahead, as model capabilities advance and multimodal integration deepens, generative AI is poised to deliver more profound intelligent support across an expanded range of structured and unstructured tasks.

### 2.2 Multimodal Retrieval and Knowledge Graph Integration

The rapid advancement of artificial intelligence (AI)—particularly the deep integration of multimodal learning and knowledge graphs—is fundamentally reshaping traditional information retrieval. Modern libraries curate increasingly diverse collections, encompassing not only conventional textual resources but also large volumes of images, audio, video, and structured data as heterogeneous information carriers. Powered by state-of-the-art models such as DeepSeek and LayoutLMv3, retrieval systems can achieve unified semantic understanding of cross-modal data and enable correlated cross-modal retrieval. For example, these models can accurately parse text and layout structures in ancient documents, extract key information from academic charts and graphs, and interpret audio-visual content in videos—ultimately facilitating the construction of cross-modal knowledge graphs that integrate heterogeneous information [4,5].

Such knowledge graphs not only unify data from diverse sources and formats into a cohesive knowledge network but also establish rich semantic connections through entity recognition, relation extraction, and graph neural network (GNN) technologies. As a result, users can retrieve semantically relevant multimodal resources using natural language queries, a single image, or an audio clip—completely breaking down the traditional medium-specific barriers in retrieval. For instance, uploading a historical map prompts the system to recommend related archives, research papers, and even digital exhibition resources; providing a lecture recording enables the system to locate corresponding lecture notes and slides.

This integrated retrieval mechanism significantly expands the breadth and depth of information retrieval, enabling libraries to deliver more comprehensive, multi-dimensional, and efficient knowledge discovery experiences for users. It represents a genuine service transformation from "retrieving materials" to "deriving insights."

### 2.3 Personalized Recommendation and User Profiling

AI systems, powered by machine learning and deep learning algorithms, can conduct in-depth analysis of multi-dimensional user data—including borrowing history, search behaviors, browsing trajectories, download records, and even page dwell time—to build dynamic, precise user interest profiles. Unlike traditional rule-based recommendation methods or simple collaborative filtering, modern AI recommendation models such as BiLSTM-CRF and LightFM capture not only users' explicit preferences but also their latent research interests and cross-domain knowledge needs, enabling truly personalized resource recommendations [6]. The large model-based personalized resource recommendation system delivers convenient, accurate, and tailored academic resource recommendations through the synergetic operation of three core modules: precise user profile construction, resource semantic index generation, and intelligent resource recommendation—as illustrated in Figure 1.

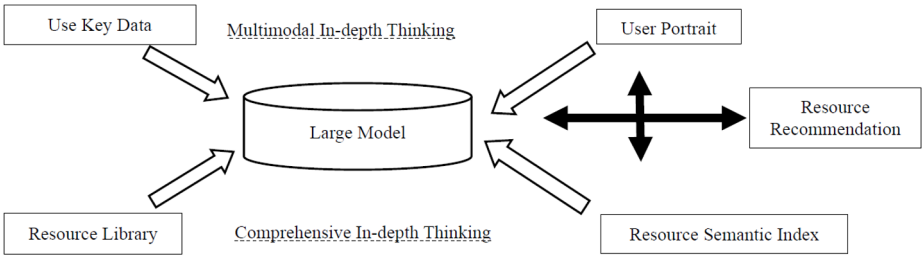


**Fig. 1.** Flowchart of Personalized Resource Recommendation System Based on Large Model

Take academic libraries as an example: these systems use sequence modeling to analyze the evolution of readers' research topics over time, employ attention mechanisms to distinguish between long-term and short-term interests, and deeply integrate document semantic content with user behaviors via knowledge graphs. For instance, when a user continuously searches for literature related to "the application of machine learning in remote sensing images", the system can not only recommend the latest high-quality papers and technical reports but also further provide relevant academic conference information, cutting-edge review articles, and even open datasets. This enables a service transformation from "passive response" to "active empowerment", and from "literature searching" to "solution provision", ultimately generating an intelligent knowledge service package as illustrated in Table 1.

**Table 1.** Intelligent knowledge service scheme table

| Recommendation Type | Content Examples | Service Value |
|---|---|---|
| Latest Paper Recommendation | <<Transformer-based Models for Agricultural Image Segmentation>>(202X) | Provides references for cutting-edge methods |
| Academic Conference Reminder | Submission Deadline Notice of IGARSS 202X+Collection of Outstanding Papers from Previous Sessions | Facilitates paper publication |
| Open Dataset Link | EuroSAT Download Link + Sample Preprocessing Script | Lowers the threshold for experiments |
| Related Book Recommendation | <<Practical Guide to Python Remote Sensing Image Processing>>, <<Fusion of Deep Learning and GIS>> | Supplements tool operation skills |
| Scholar Community Suggestion | Recommendation to join the "Young Scholars Exchange Group for Agricultural AI" | Promotes collaborative communication |

Such personalized services have significantly improved resource discovery efficiency and user satisfaction, while also facilitating interdisciplinary knowledge exploration and integration. In doing so, they transform libraries into intelligent knowledge hubs that underpin academic innovation and lifelong learning.

### 2.4 Intelligent Q&A and Reference Consultation

Conversational AI robots, powered by generative AI and semantic understanding technologies, are emerging as a core component of next-generation library reference consultation service systems. Typical applications such as Scopus AI and Web of Science Assistant provide users with round-the-clock, real-time intelligent Q&A and reference consultation services [7]. Unlike traditional systems based on keyword matching or fixed Q&A databases, these AI assistants possess advanced intent recognition and multi-turn dialogue management capabilities: they can accurately interpret complex natural language queries, dynamically supplement, clarify, and refine question details during interactions, and deliver a truly human-like interactive experience.

For example, a user might pose a composite query: "Please help me find highly cited review papers on 'artificial intelligence ethics' from the past three years, and compare the differences in research positions between the EU and China." The system can decompose multiple sub-intents, locate relevant literature, generate comparative summaries, recommend key authors or institutions, and respond to follow-up questions in dialogue—delivering highly structured, contextualized answers.

These systems not only alleviate the repetitive consultation burden on professional librarians but also drastically improve the accessibility, accuracy, and response efficiency of information services. This makes high-end, personalized reference consultation no longer constrained by time or human resources, driving the in-depth development of library services toward intellectualization and inclusiveness. Combined with the application of intelligent Q&A and reference consultation systems in university libraries, this paper conducts a data-driven comparative analysis between the traditional reference consultation mode and the intelligent Q&A system based on generative artificial intelligence from multiple dimensions, including service efficiency, user satisfaction, resource utilization rate, and librarians' workload. It quantitatively demonstrates the transformative improvement brought by technological empowerment. A comparative analysis of the intelligent question answering system and traditional reference service is presented in Table 2.

**Table 2.** Data comparison table between intelligent q&a system and traditional reference consultation

| Evaluation Dimension | Traditional Reference Consultation Model | Intelligent Q&A & AI Reference System (e.g., Large Model + Knowledge Graph) | Improvement & Key Insights |
|---|---|---|---|
| Average Re- | 4–24 hours (de- | < 30 seconds | >95% reduction |

| | | | |
|---|---|---|---|
| sponse Time | pendent on librarians' working hours) | (7×24-hour real-time response) | in latency; enables instant service access |
| Service Accessibility (Time Coverage) | 8:00–17:00 on workdays; closed on holidays | 7×24-hour online service without interruption | >300% expansion of service window; eliminates temporal barriers |
| Complex Problem-Solving Ability | Relies on librarians' expertise; requires manual retrieval across multiple databases | Automatically decomposes composite questions (e.g., "highly cited reviews + comparative research stances"); integrates multi-source knowledge bases and semantic analysis | Supports multi-turn dialogue clarification and intent tracking; accuracy improved to >85% (test data) |
| User Problem Resolution Rate | About 60–70% (limited by librarians' knowledge breadth and time constraints) | 80–92% (based on test samples, including literature retrieval, citation analysis, and trend comparison) | 15–25 percentage points increase in effectiveness |
| Librarians' Repetitive Workload | High-frequency questions account for 60–70% of total consultations | Automatically handles over 80% of common inquiries; librarians focus on high-value services | Reduces repetitive tasks by 2/3 |
| High-Level Service Conversion Rate | Users only obtain basic answers; minimal extended exploration | Proactively recommends relevant literature, research trend maps, and academic communities; 35% of users initiate secondary in-depth consultations | Enables a closed-loop workflow: Q&A → Discovery → Research |
| Concurrent Processing Capacity | 1–2 consultations per librarian simultaneously | Supports concurrent responses to thousands of users without queuing delays | Scalable to large-scale user groups |

# 3 Challenges Faced by Ai-Enabled Information Retrieval

## 3.1 Data Security and Privacy Protection

While artificial intelligence (AI) systems enhance the intelligence of information retrieval, they also pose prominent challenges to data security and privacy protection. Such systems typically rely on large volumes of user behavioral data—including search records, browsing histories, literature download preferences, and even page dwell time—to train and optimize models, thereby achieving more accurate semantic understanding and personalized services. However, the collection, storage, and processing of these data entail inherent risks of leakage, unauthorized access, and misuse. This issue is particularly acute in fields involving highly sensitive information, such as medicine, law, and finance [8].

In medical libraries, for instance, users' search content may involve disease diagnoses, medication regimens, or clinical research protocols. If such data are leaked or linked to specific individuals, it will not only violate user privacy but also trigger severe ethical and legal repercussions. Similarly, in the legal domain, search behaviors may reveal case defense strategies or sensitive litigation information, necessitating an extremely high level of security safeguards.

To address these challenges, advanced technical measures—such as federated learning, differential privacy, and encrypted computing—should be adopted to enable model training on the premise of "data not leaving the domain." At the governance level, libraries need to establish rigorous data compliance policies, clarify the ownership and usage boundaries of data, and introduce third-party audit and ethical evaluation mechanisms. This ensures that while AI applications enhance service capabilities, they consistently adhere to legal and ethical requirements for privacy protection and data security.

## 3.2 Issues of Algorithm Transparency and Bias

Although the application of AI models in information retrieval has significantly improved efficiency, it has also aroused widespread concern regarding algorithm transparency and bias. Generative models represented by ChatGPT exhibit a notable "hallucination" phenomenon [9]—that is, the model may generate content that appears plausible but contains factual errors or fabricated information. In academic search scenarios, for example, such models might produce non-existent paper titles, incorrect author information, or false experimental data, thereby misleading users and seriously undermining the rigor and reliability of academic research.

Furthermore, biases embedded in algorithm training data can lead to unfair or one-sided search results and recommended content. For instance, if social science research in the training data is overly concentrated on Euro-American perspectives, the system may overlook relevant research outcomes from Asia, Africa, and other regions when responding to user queries about "community governance models" or "changes in family structures," inadvertently reinforcing geographical imbalances in academic discourse. Similarly, when recommending influential literature or representative au-

thors, algorithms may systematically marginalize the contributions of female scholars or researchers from developing countries due to biases in training corpora.

Such issues not only compromise the accuracy and fairness of information retrieval but also risk solidifying or even amplifying existing academic biases in reality. Therefore, improving model interpretability, implementing manual or automated verification of generated results, applying bias correction technologies, and constructing more diverse and balanced training corpora have become indispensable approaches to advancing AI retrieval systems toward greater reliability and fairness.

### 3.3 Technology Adaptation and Talent Shortage

In the process of introducing AI technologies to enhance information retrieval capabilities, libraries confront significant challenges related to technology adaptation and talent scarcity. At the technical level, existing library management systems are often built on traditional architectures, which differ substantially from next-generation AI tools (such as generative models and deep learning recommendation systems) in terms of data interfaces, operating environments, and computing architectures. This results in complex system integration and low collaborative efficiency. Meanwhile, many AI models require extensive localized optimization and continuous parameter adjustment based on the characteristics of library collections and user behaviors during actual deployment. However, libraries generally lack sufficient computing resources and algorithmic engineering support, making it difficult to effectively complete model fine-tuning and ensure stable performance.

More critically, talent shortage has emerged as a key bottleneck restricting the implementation of AI in libraries. Libraries are in urgent need of interdisciplinary talents who possess proficiency in AI technologies and a deep understanding of library and information science business logic [10]. Such talents must not only master core technologies such as natural language processing, knowledge graph construction, and multimodal analysis but also demonstrate capabilities in data governance, user behavior analysis, and information service design. Nevertheless, the current university training system still largely segregates technology and humanities disciplines, leading to a severe shortage of such interdisciplinary professionals. While the existing librarian team has rich experience in literature organization and user services, most lack essential algorithmic literacy and the ability to update technical skills; conversely, technical professionals often lack in-depth insight into library business scenarios and user needs. Communication and collaboration barriers between these two groups further hinder the progress of AI application promotion.

To systematically address the challenges of technology adaptation and talent shortage, libraries need to strengthen technology selection and top-level design. Simultaneously, they should actively collaborate with universities and enterprises to establish joint training mechanisms. Additionally, libraries should gradually enhance the overall digital literacy and AI application capabilities of their teams through professional training, technology introduction, and organizational restructuring.

### 3.4 User Trust and Ethical Risks

With the widespread application of generative AI in library information retrieval, user trust and ethical risks have gradually evolved into unignorable key issues. Although AI systems can efficiently generate fluent and structured textual content, some users—especially those in scientific research and education fields—remain cautious or even skeptical about the accuracy and reliability of the outputs. This lack of trust primarily stems from the "black box" nature of models: users struggle to understand the internal logic or underlying basis of the generated answers, thereby reducing their willingness to use AI in formal and critical scenarios.

Particularly in highly sensitive contexts such as academic writing and knowledge verification, the misuse or abuse of AI technology may trigger serious academic integrity issues [11]. For example, if students or researchers directly use tools like ChatGPT to generate literature reviews, experimental analyses, or even initial drafts of papers without standardized citation and rigorous fact-checking, it may not only result in factual errors or logical flaws in the content but also cross the red line of academic plagiarism and fraud. In recent years, there have been multiple cases where academic authors have had their papers retracted or even faced integrity investigations due to the improper use of AI-generated text, reflecting the lag in the identification and management of AI-generated content within the current academic community.

**Table 3.** Comparative overview of core dimensions of ai-enabled information retrieval in libraries

| Dimension Category | Core Content | Existing Challenges | Optimization Strategies |
|---|---|---|---|
| Technology Application | Generative AI-powered intelligent Q&A, multimodal retrieval, knowledge graph integration, personalized recommendation engines, and AI-enhanced reference | Low algorithmic transparency/inherent algorithmic bias, Poor technological adaptability (poor interoperability between traditional library architectures and AI tools), inadequate system scalability for high-concurrency scenarios superscript | Promote the "Generative AI + Traditional Retrieval" hybrid model; explore applications of federated learning and edge computing for decentralized data collaboration superscript; implement dynamic routing mechanisms based on query complexity |
| Data Governance | Collection and analysis of user behavioral data, standardization of | Risks of data leakage and misuse, insufficient privacy protection measures, | Establish full-lifecycle data quality management systems; adhere to the principle of least |

| | | | |
|---|---|---|---|
| | multi-source heterogeneous data, and data infrastructure support for personalized services superscript | non-compliance with data regulatory requirements, and lack of semantic consistency in multi-source data | privilege for data access; adopt encrypted transmission, data desensitization, and third-party audit mechanisms |
| Talent System | Development of librarians' AI literacy (including prompt engineering, model fine-tuning, and human-AI collaborative workflow design), superscript provision of user education on AI information literacy | Shortage of interdisciplinary talents with both library science and AI expertise, inadequate technical competence of librarians in emerging AI tools, and weak awareness of critical AI usage among users superscript | Launch university-enterprise joint training programs; develop specialized AI training modules for librarians; construct user-centered workshop and curriculum systems on AI literacy superscript |
| Ethical Norms | Definition of liability for AI-generated content, delineation of academic integrity boundaries, and review of algorithmic fairness and transparency superscript | Erosion of user trust, academic misconduct induced by improper AI use, absence of systematic ethical risk assessment mechanisms, and ambiguity in AI accountability frameworks | Formulate clear AI usage guidelines for library services; establish regular ethical risk audit procedures; promote industry-wide collaboration in developing AI compliance and governance standards |

Furthermore, over-reliance on AI may gradually erode users' research and critical thinking abilities, which is detrimental to the healthy development of the academic ecosystem in the long run. Therefore, while introducing AI retrieval and generation tools, libraries also need to formulate supporting clear usage guidelines and ethical norms, strengthen information literacy education, guide users to utilize AI in a reasonable, transparent, and responsible manner, and gradually build a credible, controllable, and ethically sound smart service environment. The comparison of core dimensions of AI-assisted information retrieval in libraries is shown in Table 3.

# 4 Development Suggestions and Future Outlook

## 4.1 Promote the Integrated Model of "Generative AI + Traditional Retrieval"

Against the backdrop of the continuous innovation of artificial intelligence (AI) technologies, building a dual-engine retrieval system of "Generative AI + Traditional Retrieval" has become an important path for the transformation and upgrading of library information services. Generative [12]. AI excels in semantic understanding, intelligent Q&A, and personalized recommendation, which can significantly enhance the intelligence level of retrieval and user experience. In contrast, traditional retrieval methods based on keyword matching and Boolean logic, despite their limited flexibility, have advantages such as stable results, strong interpretability, and high controllability. The organic integration of the two can not only give full play to generative AI's capabilities in natural interaction and multi-dimensional correlation but also rely on traditional retrieval to ensure the precision ratio of core literature and system reliability, forming a new-generation retrieval architecture with complementary functions and synergistic effects.

In practical applications, the system can first use generative models to perform semantic parsing and intent recognition on user queries, expand retrieval keywords, construct retrieval strategies, and initially screen relevant literature. Then, it can leverage traditional retrieval methods to conduct precise matching and Boolean filtering on core databases to ensure high relevance and authority of results. For example, when responding to the query "the impact of global climate change on the agricultural economy", the system can first generate relevant subject terms and retrieval formulas, then use traditional retrieval to ensure the comprehensive acquisition of high-level journal literature, and finally utilize AI to summarize results and provide intelligent interpretation.

In the future, as the hybrid model matures, libraries can further explore a dynamic routing mechanism for the retrieval process—i.e., automatically assigning retrieval paths based on query complexity, user preferences, and resource types. This achieves an intelligent balance between flexibility and controllability, providing users with an efficient and reliable knowledge discovery experience.

## 4.2 Strengthen Data Governance and Ethical Norms

With the in-depth application of artificial intelligence (AI) in library information retrieval, establishing a sound system of data governance and ethical norms has become an important foundation for ensuring the sustainable development and credible application of the technology [13]. At the data level, libraries need to build a systematic data quality management framework covering the entire lifecycle of data collection, cleaning, annotation, storage, and usage. Special attention should be paid to the standardization and semantic consistency of multi-source heterogeneous data to provide a reliable data foundation for AI model training and optimization. Meanwhile, in response to users' privacy and security needs, it is imperative to strengthen encrypted transmission and storage, fine-grained access control, and data desensitization mech-

anisms. Particularly when processing sensitive information such as borrowing history, search behavior, and personal preferences, the "principle of least privilege" must be strictly followed to prevent risks of data leakage and misuse.

In terms of ethical norms, libraries should take the lead in formulating clear guidelines and operational standards for AI usage, including the attribution of responsibility for generated content, boundaries of academic integrity, requirements for algorithm transparency, and mechanisms for user informed consent. For example, clear regulations can be established on the scope and methods of AI tool usage in scenarios such as reference consultation and academic writing assistance to avoid academic misconduct caused by technology abuse. In addition, an ethical risk assessment and audit mechanism for AI systems should be established to conduct regular reviews of algorithm bias, fairness, and interpretability, ensuring that their decision-making processes and outcomes comply with ethical requirements and social values.

In the future, libraries also need to actively participate in the construction of industry consensus, promote the formation of standards for intellectual property, privacy protection, and ethical governance adapted to AI development, so that technological progress and normative safeguards evolve in tandem. Ultimately, this will realize the compliant, responsible, and human-centric implementation of AI in information services.

### 4.3 Improve Librarians' AI Literacy and Strengthen User Education

Against the backdrop of AI-driven changes in information retrieval, improving librarians' AI literacy and strengthening user education have become key tasks for libraries to adapt to the transformation toward smart services. As core implementers of information services, librarians need to gradually master the necessary AI technology application capabilities. Libraries should systematically launch training programs for librarians, covering aspects such as the basic principles of generative AI, prompt engineering, model fine-tuning, and the design of human-AI collaborative workflows [14]. For example, through learning prompt engineering, librarians can construct user queries more accurately, significantly improving retrieval efficiency and result relevance; meanwhile, model fine-tuning skills enable them to optimize localized AI service models based on the characteristics of the library’s collections and user habits, providing intelligent services that are more aligned with actual needs.

At the same time, in the face of an increasingly intelligent retrieval environment, users also need to possess corresponding cognitive and usage capabilities. Libraries can help users understand the mechanisms and limitations of AI retrieval, and cultivate their awareness of critical use of AI tools through various methods such as organizing workshops, creating promotional materials, and offering information literacy courses. For instance, launching themed workshops titled "AI Retrieval: Opportunities and Challenges" to guide users to distinguish the credibility of generated content, understand the potential "hallucination" and bias issues of AI, and master methods to verify the authenticity of information. These initiatives not only help users utilize AI retrieval tools more efficiently and safely but also prevent, to a certain extent, academic integrity issues caused by technology abuse.

Through the dual-track education strategy for librarians and users, libraries can build a healthier and more sustainable human-AI collaborative service ecosystem, ensuring that AI technology truly empowers information retrieval and drives the development of knowledge services toward greater intelligence and humanization.

### 4.4 Explore the Application of Federated Learning and Edge Computing

To address the challenges of data privacy and system performance faced by artificial intelligence (AI) in the application of library information retrieval, emerging technologies such as federated learning and edge computing provide effective solutions [15]. Federated learning enables multiple libraries to realize the collaborative utilization of data resources and joint modeling on the premise of fully protecting user privacy by completing model training locally and only sharing model parameters instead of raw data. For example, different university libraries can jointly train an inter-institutional academic resource recommendation model through a federated learning framework. This not only enriches the diversity of training samples and improves recommendation accuracy, but also strictly avoids the leakage and cross-domain flow of sensitive user data, complying with increasingly stringent data compliance requirements.

In terms of improving the performance of retrieval systems, edge computing technology significantly reduces network transmission latency and increases retrieval response speed by offloading part of the computing tasks to edge nodes close to data sources (such as local servers or regional data centers). Particularly in scenarios involving high-concurrency queries, multimodal retrieval, or real-time personalized recommendations, edge nodes can quickly complete preliminary computing and cache scheduling, effectively reducing the load pressure on central servers and enhancing system stability and scalability. For example, in mobile library apps, frequently initiated user retrieval requests can be directly responded to by edge nodes, greatly shortening waiting time and improving user experience.

The combination of federated learning and edge computing provides a solid technical foundation for building a new-generation intelligent library retrieval system that is efficient, secure, and collaborative, driving the broader intelligent upgrading of knowledge services on the premise of privacy protection.

## 5 Conclusion

Artificial intelligence (AI) technology is profoundly reshaping the ecosystem of library information retrieval, driving a paradigm shift from traditional tool-assisted retrieval to comprehensive intelligent empowerment. Applications including generative AI, cross-modal retrieval, and personalized recommendation have not only drastically enhanced the efficiency and quality of information services but also elevated the naturalness and efficiency of the user experience.

In response to this transformative shift, libraries must continue to advance exploration across multiple dimensions: technology integration, data security, talent development, and ethical governance. On the one hand, they should proactively promote

the "Generative AI + Traditional Retrieval" hybrid model to strike a balance between intelligence and controllability; on the other hand, it is imperative to strengthen data governance and privacy protection, and establish robust ethical norms for AI utilization. Concurrently, libraries should enhance human-AI collaboration capabilities by elevating librarians' AI literacy and delivering targeted user information literacy education.

Looking ahead, libraries should endeavor to build a new paradigm of smart services driven by the dual pillars of "artificial intelligence + humanistic care". While leveraging technological empowerment to enhance services, they must uphold their core service mission—ultimately realizing the transformation from a "knowledge warehouse" to a "smart hub", and evolving into a trusted smart center that underpins scientific research innovation and public lifelong learning.

## Acknowledgments

This study was funded by Key Project of Humanities and Social Sciences Research in Universities of Anhui Provincial Department of Education in 2024 (2024AH052237), Key Academic Research Project of Anhui Library Society (2025AHXHZD04), Project of Anhui University Library and Information Work Steering Committee (TGW24B07).

## References

1. Wang Dan, Wang Youmei, et al.: "Generative" or "Traditional"?——An Empirical Study on Postgraduates' Information Retrieval Behavior in the Age of Artificial Intelligence[J]. Research on Library Science (07), 79-91+114(2025)
2. Guo Yajun, Zhou Jiahua, et al.: ChatGPT Empowers Information Retrieval: Principles, Evaluation, Scenarios and Approaches [J]. Information Studies: Theory & Application 48(01), 83-92 (2025)
3. Hou Zhijiang.: A New Mode of Knowledge Acquisition Based on the Integration of Generative AI and Information Retrieval[J]. Library Theory and Practice (04), 86-95(2025)
4. Han Cheng.: Exploration on the Application of AI in Information Retrieval in Universities[J]. China Computer & Communication 37(15), 176-178(2025)
5. Wang Xingzi.: Research on the Application of DeepSeek in the Development of Libraries[J]. New Legend (27), 114-116(2025)
6. Du Jun.: Application and Challenges of Artificial Intelligence in Personalized Recommendation Systems of University Libraries[J]. China CIO News (08), 108-110(2025)
7. Cui Liyuan, Qiu Yuhong.: Current Situation, Challenges and Enlightenments of Conversational AI Robots in Medical Information Retrieval[J]. Jiangsu Science & Technology Information 42(07), 50-54(2025)
8. Yan Zhouzhou, Luo Guofeng, Yi Tong.: A Practical Study on ChatGPT-Driven Information Literacy Education in Universities[J]. Journal of Library and Information Science Guidance,10(07), 42-49(2025)
9. An Hongjie, An Yang.: Research on Conversational Large Models Empowering Information Retrieval in Smart Libraries[J]. The Library Journal of Henan 45(04): 78-81(2025)
10. Wang Shanshan, Ruan Mei.: Research on the Empowerment of Artificial Intelligence to Libraries[C]//University Library and Information Work Steering Committee of the Minis-

try of Education, CALIS Management Center, Shanghai Jiao Tong University Library. Proceedings of the 16th Forum on Library Management and Service Innovation 2023. University of Shanghai for Science and Technology; Shanghai R&D Public Service Platform Management Center, 60(2023)
11. Liu Weiguo, Huang Xiangxi.: A Comparative Study of Artificial Intelligence and Information Retrieval[J]. Journal of Information Science (01), 51-55(1988)
12. Guo, Jiafeng; Fan, Yixing; et al.: A Deep Look into neural ranking models for information retrieval[J]. Information Processing & Management 57(6), 102067(2020)
13. Asim, Muhammad; Arif, Muhammad; et al.: Investigating applications of Artificial Intelligence in university libraries of Pakistan: An empirical study[J]. The Journal of Academic Librarianship 49 (6), 102803(2023)
14. Luo Guofeng; Liu Qingsheng.: Research on Application Scenarios and Practices of ChatGPT Empowering Information Literacy Education in Universities[J]. Journal of Library and Information Sciences in Agriculture 36(4), 91-101(2024)
15. Siragusa, Irene; Contino, Salvatore; et al.: MedPix 2.0: A Comprehensive Multimodal Biomedical Data Set for Advanced AI Applications with Retrieval Augmented Generation and Knowledge Graphs[J]. Data Science and Engineering 5, 02994(2025)