

Research on the Ethical Governance Framework for AI Applications in Libraries

Shan Xu Zhang Chengyu

2026-07-24

ChinaRxiv

View the original and related papers at
<https://chinaxiv.org/items/chinaxiv-202607.00241>

Source: ChinaXiv. Machine translation. Verify with the original.

Abstract

The application of AI technologies in the library field has become increasingly extensive, improving service efficiency while also bringing numerous ethical challenges. This article deconstructs the 14 key issues proposed in the IFLA statement “Libraries and Artificial Intelligence,” categorizing them into three dimensions: technological risks, social risks, and governance risks, and analyzes the application limitations of this guideline as shaped by a Western context. It also focuses on the policy adaptability of AI ethics governance in Chinese libraries, with particular attention to the points of alignment between the Interim Measures for the Management of Generative Artificial Intelligence Services and this guideline, and systematically examines multiple pain points in current practice. On this basis, the article constructs a localized AI ethics governance framework encompassing five core dimensions: governance principles, governance objects, governance subjects, governance content, and governance tools, and designs a dynamic evaluation mechanism combining quantitative and qualitative approaches.

Full Text

【Theory • Exploration】

Research on the Ethical Governance Framework for AI Applications in Libraries

—An Exploration of Chinese Practice Based on the IFLA Guidelines

Shan Xu Zhang Chengyu
Tsinghua University Library, Beijing, 100084

Abstract The application of AI technology in the library field is becoming increasingly widespread. While improving service efficiency, it has also brought numerous ethical challenges. This article deconstructs the 14 key questions proposed in IFLA’s “*Libraries and Artificial Intelligence*” *Entry Points*, classifies them into three major dimensions—technical risks, social risks, and governance risks—and analyzes the application limitations of this guideline as formed within a Western context. At the same time, focusing on the policy adaptability of AI ethical governance in Chinese libraries, the article examines in depth the points of convergence between the *Interim Measures for the Management of Generative Artificial Intelligence Services* and the guideline, and systematically analyzes the multiple pain points in current practice. On this basis, the article constructs a localized AI ethical governance framework covering five core dimensions: governance principles, governance objects, governance actors, governance content, and governance tools; it also designs a dynamic evaluation mechanism combining quantitative and qualitative approaches.

[Keywords] AI Library Ethical governance IFLA

[Chinese Library Classification No.] G250.7 [Document Code]
A [Article No.] 1003-7845(2026)02-0007-08

[Citation Format] Shan Xu, Zhang Chengyu. Research on the Ethical Governance Framework for AI Applications in Libraries—An Exploration of Chinese Practice Based on the IFLA Guidelines[J]. *University Library Work*, 2026(2): 7-14.

Introduction

The rapid development of AI technology is penetrating all sectors of society at an unprecedented speed and breadth. As important institutions for information and knowledge services, libraries are also undergoing profound changes brought about by AI technology. As Chen Ying^[1] points out in her study, generative AI is a “double-edged sword” for libraries: it provides important opportunities for libraries to improve service efficiency and optimize user experience, while also exposing them to ethical dilemmas such as insufficient information accuracy, algorithmic bias, privacy leakage, and intellectual property disputes. How to effectively avoid and manage the potential risks of AI while embracing the convenience brought by the technology, and how to ensure the responsible application of AI in the library field, have become urgent issues currently facing the library community.

As an authoritative organization in the global library community, the International Federation of Library Associations and Institutions (IFLA) pays close attention to the development of AI technology and its application in libraries, and actively explores practical pathways for the standardized application of AI technology. IFLA has issued a series of guidelines intended to guide the global library community in upholding core values, safeguarding intellectual freedom, and promoting the ethical application of AI technology in the AI era. Among them, IFLA’s “*Libraries and Artificial Intelligence*” *Entry Points* (hereinafter referred to as the *Guidelines*) explicitly proposes a checklist of issues that librarians should focus on when assessing and discussing the benefits and risks of AI, providing important framework-based guidance for the ethical governance of AI in libraries^[2]. However, when these guidelines are implemented in concrete practice, it is necessary to fully consider the laws and regulations, cultural backgrounds, and practical contexts of different countries and regions. In China, with the promulgation of the *Interim Measures for the Management of Generative Artificial Intelligence Services* (hereinafter referred to as the *Measures*), how to integrate IFLA’s AI ethical framework with China’s domestic policies and regulations as well as industry practice, and how to construct a library AI ethical governance system suited to China’s national conditions, have become key concerns in the current library community. Therefore, taking the *Guidelines* as a reference and combining them with China’s specific context, this article conducts research on localized adaptation and explores in depth the construction

path of a library AI ethical governance framework, with a view to promoting the sustainable development of AI technology in the library field.

1 Deconstruction of the IFLA Ethical Framework

The *Guidelines* elaborate on the potential risks of applying AI technology in the library field and propose 14 key questions that librarians should focus on when evaluating AI services. These questions constitute the core of the ethical framework for AI in libraries and are intended to guide the global library community in applying AI technology responsibly. To systematically understand and respond to the above risks, this article classifies the 14 key questions into three major dimensions: technical risks, social risks, and governance risks, so as to more accurately grasp IFLA's core considerations in the field of library AI ethics.

1.1 Three-Dimensional Classification of the 14 Key Questions

1.1.1 Technical Risks Technical risks mainly concern the defects and limitations of AI systems themselves

Author profiles: Shan Xu, master's degree, engineer; research interests: smart libraries and knowledge management. Zhang Chengyu, PhD, associate research librarian; research interests: smart libraries and knowledge management.

Date received: 2025-09-15 Responsible editor: Jia Nan

and the problems that may arise in the process of information processing. The following issues raised in the *Guide* are closely related to this dimension: Inaccurate and misleading information (Issue 2). AI operates on the basis of probabilistic computation and may generate erroneous information, outdated information, and fabricated citations. This is directly related to the library's core responsibility as an information provider—ensuring the accuracy and reliability of information. When AI-generated content contains “hallucinations” or false information, it will seriously damage the public credibility of libraries^[3]. Lack of explainability and transparency (Issue 7). The internal operating mechanisms of AI are often opaque; the use of training data is not disclosed, and responsibility is unclear when AI makes errors, all of which constitute risks at the technical level. Libraries need to understand AI decision-making processes so that they can explain them to users and assume responsibility^[4]. Exaggerated claims and hype regarding proven benefits (Issue 1). Excessive market promotion of AI may lead libraries to blindly follow trends in technology selection and to overlook simpler, effective, and fairer non-AI solutions. This requires libraries to conduct prudent assessments of the actual benefits of AI and avoid being misled by technological hype^[5].

1.1.2 Social Risks

Social risks focus on the possible negative impacts of AI technology on individuals, groups, and social structures, especially issues related to fairness, employment, and cultural diversity. The following issues raised in the *Guide* are closely related to this dimension: Bias (Issue 3). In addition to bias caused by improper algorithm design, AI outputs may reflect uncontrolled bias present in training data, which runs counter to the library value of promoting equitable access to information^[6]. Threats to respect for cultural and linguistic diversity and different knowledge systems (Issue 4). Developing AI in a non-inclusive way may reproduce colonial power structures and thereby weaken libraries' efforts in decolonizing collections and practices. AI should support, rather than suppress, cultural and linguistic diversity^[7]. Inequality of access opportunities (Issue 6). The high cost of AI systems and their requirements for digital skills place users with limited digital access, insufficient relevant skills, or inadequate confidence at a disadvantage. Libraries need to ensure the accessibility and inclusiveness of AI services^[8]. Threats to human autonomy (Issue 11). Overreliance on AI may lead to the degradation of users' skills and a loss of confidence, thereby triggering excessive dependence on technology. Libraries should ensure that AI enhances rather than replaces human agency^[9]. The risk of replacing jobs or exacerbating exploitative labor relations (Issue 12). AI may replace some of the work of librarians, or lead to low-paid and unstable labor relations in such links as data annotation. Libraries need to pay attention to the impact of AI on staff employment and career development^[10]. The misuse of AI to create false information, misleading information, and censorship (Issue 5). Malicious actors may use AI technology to generate and disseminate false information on a large scale, and may even engage in information censorship, causing serious damage to information culture. As a key site for information-literacy education, libraries need to remain alert to and respond to such risks^[8].

1.1.3 Governance Risks

Governance risks concern issues relating to AI technology in legal, policy, ethical-framework, and accountability-mechanism terms, especially challenges related to copyright, privacy, environmental responsibility, and concentration of power. This dimension focuses on the absence or imperfection of institutions, rules, and power structures. Its core is an "imbalance between rules and power," rather than the direct social consequences brought about by technological applications. The relevant issues raised in the *Guide* include: Threats to privacy and security (Issue 8). In the current model of AI development, there is a trend of extracting user data for AI training without users' explicit consent or without providing opt-out options. As important stewards of user data, libraries must ensure data security and privacy protection^[11]. Threats to copyright and authors' rights (Issue 9). The AI training process may use copyright-protected materials, involving problems of illegal access and potential hidden costs, which has aroused concerns across sectors about the protection of intellectual prop-

erty and the rights and interests of creators^[12]. Lack of consultation with stakeholders (Issue 10). The AI development process often fails to fully incorporate and solicit the views of the broadest range of stakeholders, including groups affected by AI, resulting in designs that may not fully take social needs into account. Libraries should advocate a model of AI governance involving multiple parties^[13]. Significant environmental impacts (Issue 13). Digital technologies consume large amounts of materials and energy during equipment manufacturing and transportation, and AI training and use also involve relatively high energy demands. When introducing AI, libraries need to consider its sustainability impacts^[14]. Excessive power held by monopolistic technology companies lacking regulation, and the emergence of a harmful international AI arms race (Issue 14). Many links in the development of AI are currently concentrated in the hands of entities that lack accountability mechanisms, rather than serving the interests of all humanity. Libraries should remain alert to the potential risks of technological monopoly and advocate open, non-proprietary models of AI development^[15].

1.2 Application Limitations of Scenario-Based Tools

The *Guide* and the 14 key questions it proposes for librarians to evaluate AI services provide valuable guidance for the global library community. However, the scenario-based tools it mentions were mainly formed in a Western context and may have certain limitations when adapted to China's policy environment.

- (1) Differences in policies, laws, and regulations. The *Guide* is universal in nature, but there are significant differences among countries in laws and regulations concerning AI ethics and governance. For example, China has issued the *Measures*, which put forward explicit regulatory requirements for the research and development, application, and services of generative AI, including content compliance, data security, and algorithm filing. These specific provisions may go beyond the scope of the *Guide*, or may be more binding in certain respects. Therefore, directly applying the framework of the *Guide* may not fully satisfy China's local compliance requirements.
- (2) Differences in cultural and social backgrounds. Western societies' understanding and emphasis regarding concepts such as privacy, individual rights, and freedom of speech may differ from those in China. For example, in terms of data use and privacy protection...

In this respect, China has its own distinctive sociocultural background and regulatory practices. The emphasis in the *Guidelines* on "stakeholder consultation" may, in the Chinese context, need to be adjusted in combination with local models of social governance and mechanisms for public participation.

- (3) Differences in stages of technological development and application scenarios. Although AI technology is developing rapidly worldwide, different countries and regions still differ in their investment in AI research and

development, the extent of application uptake, and specific application scenarios. The *Guidelines* may be based more on the practical experience of libraries in Western countries, whereas Chinese libraries may face different technical challenges and business needs in applying AI. For example, in specific scenarios such as AI-based metadata annotation and reference services, Chinese libraries may require more operationally actionable localized guidance.

- (4) Differences in the maturity of ethical governance systems. Western countries started earlier in the field of AI ethical governance and have accumulated more related research and practice; mechanisms such as ethics committees and industry self-regulatory organizations are relatively mature. China, by contrast, is still in a stage of rapid development and improvement in AI ethical governance, and needs to gradually build and strengthen a multi-level ethical governance system in light of its own national conditions. The *Guidelines* provide more of a principled direction; however, with regard to specific governance paths and implementation rules, the Chinese library community still needs to carry out in-depth exploration and localized innovation.

2 Current Policies and Practices for AI Ethical Governance in Chinese Libraries

With the rapid development of AI technology, the Chinese government has attached great importance to the ethical governance and standardized development of AI. The promulgation of a series of policies and regulations has provided macro-level guidance and legal safeguards for the application of AI across various industries. As important public cultural service institutions, libraries are gradually advancing AI ethical governance against this broader background. This section discusses the policy adaptability of AI ethical governance in Chinese libraries and systematically analyzes the pain points that exist in current practice.

2.1 Policy Adaptation—Points of Convergence Between the *Measures* and the *Guidelines*

China has already established a policy system with Chinese characteristics in the field of AI ethical governance, among which the *Measures* constitute an important regulation guiding the healthy development of generative AI. The *Measures* and the *Guidelines* have many points of convergence in their core concepts and specific requirements, providing a policy basis and direction for action for Chinese libraries in carrying out AI ethical governance[16].

- (1) A shared value orientation. The principle of responsibility is reflected in the *Measures*, which emphasize that providers of generative AI services should adhere to the core socialist values, ensure that generated content embodies the core socialist values, and require effective measures to pre-

vent generated content from containing false information, ethnic discrimination, gender discrimination, and the like. This is highly consistent with the core ethical principles emphasized in the *Guidelines*, such as fairness, accuracy, and avoidance of bias; both aim to promote the responsible development of AI and prevent it from causing negative impacts on society[17].

- (2) The people-centered principle. The *Measures* require providers of generative AI services to respect the lawful rights and interests of others, not to endanger others' physical or mental health, and not to infringe upon portrait rights, reputation rights, honor rights, privacy rights, and so forth. The *Guidelines* also repeatedly mention issues such as threats posed by AI to human agency and safety threats, and emphasize ensuring that users retain agency in their interactions with AI. Both place human dignity, rights, and well-being at the core of AI development.
- (3) The principle of transparency and controllability. The *Measures* require providers of generative AI services to be responsible for the legality of the sources of training data and to label generated content. The *Guidelines* point out that a lack of explainability and transparency is one of the risks of AI, and raise questions such as whether applications possess transparency in a social sense. This shows that both recognize the importance of AI transparency and are committed to enhancing the explainability and controllability of AI through technical and managerial means[18].
- (4) Security and privacy protection. The *Measures* put forward explicit requirements for both personal information protection and data security. The *Guidelines* also pay close attention to privacy and security threats, emphasizing user data management to ensure privacy and security. This reflects the shared concern of both regarding data security and personal privacy protection in the AI era.
- (5) Complementarity and linkage in specific requirements. The *Measures* provide more detailed provisions at certain operational levels and can serve as an effective supplement to the *Guidelines*, guiding AI ethical practice in Chinese libraries. In terms of content compliance, the *Measures* put forward specific requirements for the legality, authenticity, and accuracy of generated content, and clearly define mechanisms for handling illegal content, thereby providing clear review standards when libraries introduce AI-assisted content generation, such as intelligent abstracts and recommendation statements. In terms of data management, the *Measures* regulate the sources, legality, and processing methods of training data, helping libraries better comply with data ethics and legal requirements when using collection data to train AI models, and avoid infringing copyright and personal privacy. In terms of algorithm filing and security assessment, the *Measures* require generative AI services with public-opinion attributes or social mobilization capabilities to complete security assessments and algorithm filing. Although library AI applications may not directly involve

such attributes, this requirement reflects anticipation of AI risks and prior intervention, and provides libraries with a risk-assessment approach when selecting and deploying AI systems.

Therefore, when conducting AI ethical governance, Chinese libraries can take the *Guidelines* as the theoretical framework and, in combination with the specific requirements of domestic regulations such as the *Measures*, construct an ethical governance system that both aligns with advanced international concepts and is adapted to China's national conditions.

2.2 Analysis of Practical Pain Points

Although China has already established a framework for AI ethical governance at the policy level, libraries still face multiple pain points in concrete practice, and these pain points show significant differences across libraries in different regions and at different levels of development.

2.2.1 Insufficient Ethical Awareness and Professional Knowledge

Many librarians lack sys-

a systematic understanding, making it difficult to accurately identify and assess such problems as algorithmic bias, privacy leakage, and copyright infringement that may arise in AI applications. Relevant studies show that university librarians' understanding of AI concepts and applications remains only at an intermediate level; they have limited hands-on experience, and there are significant gaps in their discussion of ethical issues and participation in AI projects[19].

2.2.2 Lack of Unified Evaluation Standards and Tools

Although the *Guidelines* provide a list of 14 issues, they have not yet formed ethical evaluation standards and tools that are tailored to the specific application scenarios of Chinese libraries and that can be directly implemented. When introducing AI, libraries often rely on vendors' commitments or on their own limited judgment, making it difficult to conduct comprehensive and in-depth ethical risk assessments. This problem is particularly prominent in libraries in the central and western regions, which lack sufficient technical support and professional guidance to build their own evaluation systems.

2.2.3 Increasingly Prominent Data-Ethics Challenges

Libraries possess massive amounts of user data and collection data. How to use these data legally and compliantly for AI training while protecting user privacy and intellectual property has become a serious ethical challenge. In particular, there is a lack of clear guidance and technical support in areas such as data desensitization, anonymization, and data sharing. Although large libraries have relatively strong technical capabilities, they still face complex data-ethics issues; small and medium-sized libraries, by contrast, may encounter compliance risks in data processing because of insufficient professional capacity.

2.2.4 Difficulties in Identifying and Intervening in Algorithmic Bias

The “black box” nature of AI algorithms makes it difficult for libraries to identify and correct potential biases within them. Even when libraries become aware of the existence of bias, they lack effective technical means and intervention strategies. This issue poses challenges for all types of libraries, but libraries in developed regions may be able to obtain more professional technical support, whereas those in less-developed regions face greater difficulties.

2.2.5 Insufficient Human Resources and Funding

Ethics governance requires professional personnel and sustained financial investment, including establishing ethics review committees, conducting ethics training, and developing ethics assessment tools. This creates tremendous pressure for libraries with limited resources. Although libraries in developed regions have relatively abundant resources, they still face a shortage of professional ethics personnel; libraries in the central and western regions suffer serious deficiencies in both funding and talent.

2.2.6 Incomplete Governance Mechanisms

At present, China’s library community lacks a unified AI ethics governance mechanism and coordination platform. Libraries largely act independently, making it difficult to form joint governance capacity. At the industry level, there is neither systematic guidance and support nor channels for sharing experience and exchanging cases. This situation not only causes waste of resources and redundant construction, but also leaves some libraries without effective avenues for seeking assistance when confronted with complex ethical issues[20].

The above practical pain points indicate that Chinese libraries still have a long way to go in AI ethics governance. Relying solely on macro-level policy guidance is far from sufficient. More operationally practical guidance plans, richer training resources, and more effective support mechanisms must be provided for libraries of different types and at different levels of development, so as to bridge regional gaps and comprehensively improve the industry’s level of AI ethics governance.

3 Design of an AI Ethics Governance Framework for Chinese Libraries

In view of the universal characteristics of the *Guidelines* and the particularities of the Chinese context, this paper, drawing on the multi-level theory of AI ethics governance proposed by Mäntymäki M et al.[21], constructs a localized AI ethics governance framework for libraries that integrates the concepts of the *Guidelines* with the practical needs of China. The framework aims to build a systematic, dynamic, and operable ethical governance system, providing support

for effectively addressing various ethical challenges arising from the application of AI technologies in libraries.

3.1 Dimensions for Constructing the Governance Framework

Based on existing theoretical achievements concerning AI ethics governance frameworks, the AI ethics governance framework for libraries constructed in this paper contains five core dimensions: governance principles, governance subjects, governance objects, governance content, and governance tools. These five dimensions constitute a complete logical chain of “why to govern (principles) → who governs (subjects) → what to govern (objects) → which aspects to govern (content) → how to govern (tools).” This both reflects the macro-structure of ethics governance and incorporates the differentiated characteristics of the library industry.

Governance principles are the value foundation and fundamental norms of the framework. In light of the core concepts of the *Guidelines* and the requirements of China’s AI governance policies, six basic principles are established: people-centeredness, fairness and inclusiveness, transparency and controllability, safety and trustworthiness, compliance with laws and regulations, and sustainable development. These principles run through the entire process of all governance activities, providing action guidance for governance subjects and value benchmarks for the application of governance content and governance tools. Governance subjects constitute the organizational guarantee of the framework. A three-level collaborative governance system at the national, provincial, and institutional levels is to be built, with the division of responsibilities and coordination mechanisms of subjects at all levels clarified: at the national level, the Library Society of China takes the lead and is responsible for top-level design and standard-setting; at the provincial level, regional ethics review committees are established to undertake policy refinement and project review responsibilities; at the institutional level, individual libraries establish internal governance mechanisms and are responsible for concrete implementation and daily supervision. Governance objects define the framework’s scope of action and the targets of governance. They cover the full life cycle of library AI applications, including AI systems and technologies, data and algorithms, services and processes, personnel and users. Subsequent governance content and governance tools will be developed around these specific objects to ensure the comprehensiveness and systematic nature of ethics governance. Governance content comprises the core requirements and specific norms of the framework. Based on the three-dimensional risk classification proposed above, corresponding governance measures and control requirements are formulated for technical risks, social risks, and governance risks, forming a full-process governance content system that covers risk identification, assessment, prevention, response, and supervision. Governance tools are the framework’s means of implementation.

stages and supporting safeguards. They include three major categories—assessment tools, institutional tools, and technical tools—providing diversified meth-

ods and pathways for ethical governance. Through the systematic application of these tools, the operability and effectiveness of the framework are ensured.

3.2 Governance Principles and Governance Objects

3.2.1 Governance Principles Governance principles are the soul of the framework, providing value orientation and fundamental guidance for all governance activities. The six major principles established in this paper are as follows: People-centered. The development and application of AI should center on serving users and librarians, respect human dignity, lawful rights, and autonomy, ensure that AI always remains under human control, and strive to enhance rather than weaken human agency. Fairness and inclusiveness. Ensure the accessibility of AI services and avoid exacerbating the digital divide due to technological, economic, regional, and other factors; actively identify and correct algorithmic bias, and promote equitable access to information resources and balanced development in service quality. Transparency and controllability. Improve the interpretability and traceability of AI decision-making processes, require clear labeling of AI-generated content, protect users' right to know, and at the same time ensure that humans retain final control in key decision-making stages. Safety and trustworthiness. Establish and improve mechanisms for data security and privacy protection, strengthen the robustness and reliability of AI systems, prevent malicious attacks and data leakage, and maintain the foundation of users' trust in library AI services. Compliance with laws and regulations. Strictly abide by national laws and regulations in such fields as AI, cybersecurity, data security, and intellectual property, and integrate compliance requirements throughout the entire life cycle of AI applications. Sustainable development. Attend to the environmental impact of AI technologies, advocate the adoption of green and low-carbon AI technologies and infrastructure, and ensure that AI applications remain consistent with libraries' long-term social responsibilities and ecological goals.

3.2.2 Governance Objects Governance objects define the boundaries and specific points of action for ethical governance, and are an important prerequisite for achieving precise governance. The governance objects of this framework cover the entire life cycle of library AI applications, mainly including: AI systems and technologies. These include the various AI models, algorithms, software platforms, and related technological infrastructure used; governance must run through the full process of their design, development, procurement, deployment, operation, and retirement. Data and algorithms. These encompass collection data and user behavior data used to train AI models, as well as the design, optimization, and operation of algorithms, with emphasis on the legality, quality, and potential bias of data, as well as the fairness and transparency of algorithms. Services and processes. These refer to library business services that integrate AI technologies—such as intelligent reference, knowledge organization, and personalized recommendation—and their related workflows; governance must ensure that service processes comply with ethical norms and

are accountable for user experience. Personnel and users. These include librarians, technology developers, managers, and all users of library AI services; governance content involves training in ethical literacy, skill transformation, user education, protection of rights and interests, and the establishment of feedback mechanisms.

3.3 Governance Actors

3.3.1 National Level At the national level, the governance actor is mainly the Library Society of China. Its core responsibilities include: formulating national white papers related to library AI ethics, providing top-level design and value guidance, and clarifying the values, basic principles, and development vision of the Chinese library community in the field of AI ethics; deeply interpreting national laws, regulations, and policy documents related to AI, and, in light of the characteristics of library operations and ethical needs, offering specific recommendations for policy alignment and implementation; based on the 14 key questions in the *Guidelines* and the actual situation of Chinese libraries, systematically sorting out and classifying potential ethical risks in library AI applications, and providing guidance methods for risk identification; collecting and promoting outstanding cases and innovative experiences in domestic library AI ethics practice, thereby providing reference examples for other libraries in the industry; and establishing a mechanism for regular updates to the white paper to ensure the timeliness and forward-looking nature of its content.

3.3.2 Provincial Level At the provincial level, regional library AI ethics review committees should be established, led by library societies or library alliances in each province, autonomous region, or municipality directly under the central government. The committee's core responsibilities include: in light of the actual conditions of the province, autonomous region, or municipality and the developmental characteristics of its libraries, formulating more operable AI ethics implementation rules and guidance opinions as supplements and extensions to the white paper issued at the national level; conducting ethical review and risk assessment of major AI projects that libraries in the region plan to introduce or develop, providing professional ethical consultation and recommendations, and ensuring that projects comply with ethical norms; organizing AI ethics training for librarians in the region to enhance their ethical awareness and risk-identification capabilities, while also promoting the application of ethical assessment tools; collecting typical cases from AI ethics practice in libraries in the region, establishing a dedicated case database, and promoting regional exchange of experience and knowledge sharing; and coordinating the handling of ethical disputes and complaints arising from library AI applications in the region, providing fair and reasonable solutions. Committee members should include people from diverse backgrounds, such as library experts, AI technology experts, legal experts, ethics experts, and user representatives, so as to ensure the comprehensiveness and impartiality of the review^[22].

3.3.3 Institutional Level At the institutional level, each library should establish an internal AI ethics governance mechanism and embed AI ethics considerations into daily workflows. The main responsibilities include: establishing an AI ethics responsibility system and clarifying the specific responsibilities of each department and position in AI ethics governance; formulating detailed AI ethics operating procedures and checklists to ensure that every AI application undergoes strict ethical scrutiny; building mechanisms for user feedback and complaint handling to respond promptly to user concerns; regularly carrying out internal AI ethics training, publicity, and education to strengthen ethical awareness among all staff; and maintaining close contact with the provincial ethics review committee, with timely reporting of major ethical issues and potential risks.

3.4 Governance Content

3.4.1 Technical Risk Governance In terms of technical risk governance, it is necessary to systematically establish mechanisms for quality control and transparency assurance. First, in response to the risks that AI-generated content may be inaccurate and misleading, regular spot checks and fact-checking procedures should be established; for example, in scenarios such as intelligent consultation, routine quality monitoring should be implemented for AI-generated answers. Second, to address the lack of interpretability and transparency in AI, when libraries procure AI systems, they should explicitly require suppliers to provide explanations of model interpretability; meanwhile, “transparency labels” should be added to AI-generated content in the user interface, the sources of training data should be clearly indicated, and a standardized transparency-labeling system should be established. Finally, to avoid technological hype and blind imitation, scientific AI technology evaluation standards should be formulated, comprehensive cost-benefit analyses should be carried out, and simple and effective solutions should be prioritized to ensure that technology applications remain pragmatic.

3.4.2 Social Risk Governance Social risk governance should be grounded in fairness, inclusiveness, and human agency, and should build a multidimensional line of defense for risk prevention and control. At the algorithmic level, bias detection and correction mechanisms should be established; for example, in library automatic classification and recommendation systems, algorithmic bias detection indicators should be used regularly to assess the balance of classifications, and biased models should be retrained in a timely manner. At the cultural level, AI systems should be ensured to support the presentation of multicultural content, avoid tendencies toward cultural homogenization, and effectively protect the rights and interests of minority groups. At the service level, attention should be focused on the digital divide; by enriching interaction methods and carrying out digital skills training, the accessibility of library services should be improved. At the human level, a balance in human-machine collaboration should be maintained, overdependence on AI should be prevented, and staff re-

assignment and training plans should be formulated to address the employment impacts brought by AI. At the same time, through information literacy education and mechanisms for refuting false information, users' rights and interests as well as social fairness should be comprehensively safeguarded.

3.4.3 Responses to Governance Risks Responses to governance risks should proceed from three aspects: institutions, rights, and ecosystems. At the institutional level, strict data protection systems should be established, the principle of data minimization should be strictly followed, and technical protection capabilities should be strengthened to effectively address privacy leakage and data security threats. At the rights level, the processes for collecting and using AI training data should be regulated, sound intellectual property protection mechanisms should be established, and the attribution of responsibility at each stage of AI application should be clarified; at the same time, multi-party consultation mechanisms should be used to ensure the full participation of stakeholders in the governance process. At the ecosystem level, the carbon footprint of the full life cycle of AI applications should be assessed, green and low-carbon AI technologies should be actively promoted to reduce environmental impacts, innovation and development in open-source technologies should be supported, technological-ecosystem diversification should be promoted, and risks of technological monopoly should be prevented.

3.5 Governance Tools

3.5.1 Evaluative Tools Evaluative tools are mainly used to measure and analyze AI ethical performance, including both quantitative and qualitative categories.

- (1) Quantitative evaluation tools mainly conduct objective measurement through data indicators. Algorithmic bias detection indicators are used to quantify bias issues in AI systems, such as identifying gender bias by analyzing the degree of association between specific genders and specific occupations in AI recommendation systems, or assessing the balance of information recommended to users from different regions and groups. Indicators for content accuracy and the rate of false-information generation are implemented through regular manual spot checks of AI-generated content or by introducing third-party tools for fact-checking. Indicators for the compliance rate of data desensitization/anonymization are used to assess the degree of compliance in libraries' handling of user data in AI applications. In addition, these also include service fairness indicators that monitor differences in the quality of AI services obtained by different user groups, system transparency indicators that assess the interpretability and information-disclosure level of AI systems, and environmental impact indicators that monitor the energy consumption and environmental costs of AI applications.
- (2) Qualitative analysis tools focus on evaluation from the perspectives of

subjective experience and in-depth understanding. Establishing an AI service user committee composed of users from different backgrounds is an effective approach, enabling regular collection of users' feedback, opinions, and complaints regarding AI services. Committee members should include ordinary readers, persons with disabilities, ethnic minority groups, and others, so as to ensure that the membership structure is comprehensively representative. User complaints should be classified and analyzed in detail, covering information quality issues, bias and discrimination, privacy leakage, operational experience, copyright issues, and so forth. In addition, case-study methods, regular user-satisfaction surveys, and focus-group discussions are also important qualitative analysis tools; they help deepen understanding of typical AI ethics incidents and reveal users' genuine feelings and expectations regarding AI services.

3.5.2 Institutional Tools Institutional tools provide mechanism-based safeguards for the practical application of evaluation results and the continuous operation of ethical governance. First, a regular reporting system should be established, requiring the user committee or relevant departments to submit AI ethical performance reports to library management on a regular basis, while also putting forward targeted improvement recommendations. Second, a problem-response mechanism should be improved: libraries should promptly adjust AI service strategies, optimize algorithmic models, strengthen user education, or collaborate with technology suppliers to solve problems in accordance with evaluation results. Third, a public oversight mechanism should be established to ensure the transparency of the evaluation process and results, encourage public participation in the oversight of ethical governance, and form positive interaction. The above mechanisms ultimately serve the goal of continuous improvement, ensuring that the AI ethical governance framework and implementation strategies are continuously refined on the basis of evaluation results.

3.5.3 Technical Tools Technical tools are key supports for implementing quantitative evaluation and carrying out part of the governance content. In addition to using AI toolkits to realize model-decision explanation and bias detection, they also cover core technical means such as data anonymization processing technologies, algorithm audit platforms, and AI-system energy-consumption monitoring tools. Libraries should, in light of their own technical capabilities and needs, select appropriate open-source or commercial technical tools and integrate them into ethical governance workflows.

4 Conclusion and Recommendations

4.1 Research Contributions and Significance

The main contributions of this paper are reflected in both theory and practice. At the theoretical level, this paper systematically deconstructs the 14 key issues in the *Guidelines*, categorizing them into three major dimensions: technical risks,

social risks, and governance risks, thereby providing a clear analytical framework for understanding and applying the *Guidelines*; it conducts an in-depth analysis of the adaptability of the *Guidelines* within China's policy environment, identifying differences in cultural background, legal environment, and stage of development; and it constructs a library AI ethics governance framework comprising five core dimensions, offering a new perspective for theoretical research on AI ethics governance.

At the practical level, this paper systematically analyzes the current state and practical pain points of AI ethics governance in Chinese libraries, laying a realistic foundation for subsequent work; it provides a concrete reference framework and implementable operational path for the library community in China to carry out AI ethics governance; and it designs an operable dynamic assessment tool, offering practical methods for libraries to evaluate and improve their AI ethics performance.

4.2 Practical Recommendations

Based on the research findings, this paper proposes the following practical recommendations: (1) Advocate preventive ethics governance. Libraries should move ethical considerations forward to the design and development stage of AI systems, rather than governing only after problems occur. Ethical review should be embedded throughout the entire life cycle of AI applications, and a multi-stakeholder mechanism for ethical consensus should be established, achieving preventive goals through ethics-by-design, whole-process review, and collaborative multi-party governance [23]. (2) Promote the joint development of an industry-wide AI ethics case-sharing repository. AI ethics issues are complex, variable, and context-dependent, and libraries of different types and scales face different ethical challenges in AI applications. It is recommended that the Library Society of China take the lead, in collaboration with provincial ethics review committees, in jointly building an "AI Ethics Case-Sharing Repository." The cases should cover different AI application scenarios and practices in different types of libraries. An open online platform should be established to facilitate libraries' submission of cases, consultation of cases, and participation in evaluation and discussion [24]. (3) Strengthen capacity building and talent cultivation. Libraries at all levels should enhance the dissemination of AI ethics knowledge and specialized training, improve librarians' ethical awareness and professional competence, and focus on bridging the capacity gap between central-western and grassroots libraries through training programs organized by provincial committees; cooperation with universities and research institutions should be strengthened to cultivate interdisciplinary talent possessing both library professional literacy and AI ethics knowledge [25]. (4) Improve institutional safeguards and standards development. Industry organizations should accelerate the formulation of standards and norms for AI ethics governance, provide unified assessment tools and operational guidelines, and establish assessment and evaluation mechanisms to incorporate AI ethics governance into library performance evaluation

systems.

4.3 Future Research Prospects

Future research should proceed in five major directions: technological evolution, international comparison, tool innovation, user participation, and long-term evaluation. In terms of technology, attention should be paid to ethical challenges arising from emerging technologies such as multimodal AI and large models, and governance frameworks should be improved in a timely manner. From an international perspective, cross-national comparative research should be carried out, policy and practical experience from other countries should be learned from, and international exchanges and cooperation should be promoted. In terms of tool development, emphasis should be placed on developing intelligent and automated ethical assessment and regulatory tools to improve governance efficiency. In terms of user participation, more effective user participation mechanisms should be explored to enhance the democratic nature and transparency of governance. In terms of long-term evaluation, tracking and assessment mechanisms should be established to systematically examine the implementation effects and social impacts of governance frameworks, providing empirical evidence for their continuous optimization.

References

- [1] Chen Ying. Opportunity or crisis: reflections on libraries' application of generative artificial intelligence [J]. *Journal of Library Science*, 2024(2): 91-94, 103.
- [2] Cox A M, De Brasdefer M. IFLA AI entry point to libraries and AI [EB/OL]. [2025-08-04]. <https://repository.ifla.org/items/f197f327-dc49-4743-bb57-0a373505da8b>.
- [3] Cox A. The ethics of AI for information professionals: eight scenarios [J]. *Journal of the Australian Library and Information Association*, 2022(3): 201-214.
- [4] Lowagie H. From bias to transparency: ethical imperatives in AI-based library cataloging [EB/OL]. [2025-08-11]. <https://repository.ifla.org/items/351270b4-31b2-4481-a67f-8412e48198bb>.
- [5] IFLA. IFLA statement on core principles on libraries and copyright [EB/OL]. [2025-08-13]. <https://repository.ifla.org/items/2a10254a-ba26-4e9b-9f95-9ca685200eed>.
- [6] Lee B C G. The “Collections as ML Data” checklist for machine learning and cultural heritage [J]. *Journal of the Association for Information Science and Technology*, 2025(2): 375-396.
- [7] ARL. Research libraries guiding principles for artificial intelligence [EB/OL]. [2025-08-04]. <https://www.arl.org/resources/research-libraries->

guiding-principles-for-artificial-intelligence/.

[8] IFLA. IFLA statement on libraries and artificial intelligence [EB/OL]. [2025-08-11]. <https://repository.ifla.org/items/8c05d706-498b-42c2-a93a-3d47f69f7646>.

[9] Cox A. Ethics scenarios of artificial intelligence for information and knowledge management and library professionals [EB/OL]. [2025-08-11]. https://orda.shef.ac.uk/articles/dataset/Ethics_Scenarios_Of_Artificial_Intelligence_For_Information_And

[10] IFLA. Artificial intelligence special interest group [EB/OL]. [2025-08-04]. <https://www.ifla.org/units/ai/>.

[11] Interim Measures for the Management of Generative Artificial Intelligence Services [EB/OL]. [2025-08-11]. http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.

[12] CARL response to the consultation on copyright in the age of generative AI [EB/OL]. [2025-08-11]. <https://www.carl-abrc.ca/wp-content/uploads/2025/08/CARL-response-to-the-Consultation-on-Copyright-in-the-Age-of-Generative-AI.pdf>.

[13] Bogaards N. Ethics and AI in the Dutch public library sector [EB/OL]. [2025-08-04]. <https://repository.ifla.org/items/0e664fd9-b523-4477-975d-6b38dcbfe8a8/full>.

[14] AI and data to improve conservation management [EB/OL]. [2025-08-04]. <https://repository.ifla.org/rest/api/core/bitstreams/ebd0a97d-7cc4-4a5e-84fc-f1f201156eeb/content>.

[15] Leclair C, Termignon L. For an ethics of personalized recommendation at the French National Library [EB/OL]. [2025-08-04]. <https://repository.ifla.org/items/f8d79679-76a9-4695-b350-14a7b7002329/full>.

[16] Standardization Administration of China; Office of the Central Cyberspace Affairs Commission; National Development and Reform Commission; Ministry of Science and Technology; Ministry of Industry and Information Technology. Notice on issuing the *Guidelines for the Construction of the National New-Generation Artificial Intelligence Standards System* [EB/OL]. [2025-08-04]. https://www.gov.cn/zhengce/zhengceku/2020-08/09/content_5533454.htm.

[17] China Academy of Information and Communications Technology; JD Exploration Research Institute. *White Paper on Trustworthy Artificial Intelligence* [EB/OL]. [2025-08-04]. <http://www.caict.ac.cn/kxyj/qwfb/bps/202107/P020210709319866413974.pdf>.

[18] Zaragoza T, Nicolas Y, Le Provost A. From text to data inside bibliographic records: entity recognition and entity linking of contributors and their roles from statements of responsibility [EB/OL]. [2025-08-04]. <https://repository.ifla.org/items/65edd46c-f668-4a73-b829-e488ff99b8a9>.

[19] Lo L S. Evaluating AI literacy in academic libraries: a survey study with a focus on U.S. employees [J]. *College & Research Libraries*, 2024(5):635-668.

- [20] Lu Kang, Pan Xiaoyu, Liu Hui, et al. Research on ethical norms for library smart services from the perspective of AI governance [J]. *Library*, 2020(6):21-28, 36.
- [21] Mäntymäki M, Minkinen M, Birkstedt T, et al. Putting AI ethics into practice: the hourglass model of organizational AI governance [J/OL]. arXiv, 2022 [2025-08-04]. <https://doi.org/10.48550/arXiv.2206.00335>.
- [22] Liu X, Yu H, Zhang H C, et al. AgentBench: evaluating LLMs as agents [J/OL]. arXiv, 2023 [2025-08-04]. <https://doi.org/10.48550/arXiv.2308.03688>.
- [23] Wu J J. Using ChatGPT to evaluate and improve library websites [EB/OL]. [2025-08-04]. <https://repository.ifla.org/items/a163c211-2e23-4327-9261-e268e672dfe4>.
- [24] Uzwyshyn R. Steps towards building library AI infrastructures: research data repositories, scholarly research ecosystems and AI scaffolding [EB/OL]. [2025-08-04]. <https://repository.ifla.org/items/4a774381-6462-4718-9092-f66526abaf63>.
- [25] IFLA. 23 resources to get up to speed on AI in 2023 [EB/OL]. [2025-08-04]. <https://www.ifla.org/g/ai/23-resources-to-get-up-to-speed-on-ai-in-2023/>.

Research on an Ethical Governance Framework for Artificial Intelligence Application in Libraries: An Exploration of Chinese Practices Based on IFLA Guideline

Shan Xu Zhang Chengyu

Tsinghua University Library, Beijing, 100084

Abstract The application of artificial intelligence technology in libraries has expanded rapidly, improving service efficiency while introducing numerous ethical challenges. The article deconstructs the fourteen key issues proposed in the *IFLA Entry Point to Libraries and AI* guideline, categorizing them into three dimensions: technological risks, social risks, and governance risks. It further analyzes the limitations of the guideline, which was formed in a Western context. The study also explores the policy alignment of AI ethical governance in Chinese libraries, with particular focus on the alignment between China's *Interim Measures for the Management of Generative Artificial Intelligence Services* and the IFLA guideline, and identifies major challenges in current practice. On this basis, the article proposes a localized ethical governance framework for AI in libraries encompassing five core dimensions: governance principles, governance objects, governing entities, governance content, and governance tools. Moreover, it develops a dynamic evaluation mechanism that integrates both quantitative and qualitative approaches.

Keywords Artificial intelligence; Library; Ethical governance; IFLA

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv. Machine translation. Verify with the original.