**Information illusion in Library AI Language Generation: Risk Avoidance Strategies from an Information Literacy Perspective**

**ABSTRACT**

The library industry has enormous development prospects and challenges due to the applications and advancement of artificial intelligence (AI). Although generative AI enhances human knowledge and talents, it produces false information influencing users’ decision-making. In the context of AI-generated language in library services, this research addresses the issue of “information illusion” and examines the causes and potential hazards of generative AI “information illusion”. From the perspective of information literacy, four strategies are proposed to avoid and mitigate the risks encountered in “information illusion” and enhance users’ critical thinking skills when navigating in AI-assisted information environments.

**KEYWORDS**

AI hallucination; Generative AI; Information literacy; Libraries; Risk avoidance

**INTRODUCTION**

Artificial Intelligence (AI) has progressed tremendously in the digital information environment, boosting the effectiveness of social services and supporting the transformation and development of services in numerous industries. In the library industry, artificial intelligence applications, incredibly generative artificial intelligence (AIGC), have become critical manifestations of library services. The application of generative AI in libraries has changed how librarians work and how libraries are managed, bringing higher quality and more comprehensive venue space services, information retrieval and reference consulting services and digital reading services. However, in the actual interaction process, the risks brought by the defects of generative AI also bring many difficulties and challenges. People who access information from generative AI to make behavioral decisions may encounter AI hallucinations. These are deceptive outputs that lead to cognitive biases and mistakes. Therefore, having high information literacy is critical to identifying and avoiding these application risks. With the continuous progress and advancements in AI technology, the significance of information literacy will become increasingly prominent. Libraries, as the primary organizations for information literacy education, should embrace this trend and offer extensive AI literacy education resources and support to the public. This paper analyzes the risk of misinformation generated by AI language models in libraries and proposes strategies for managing this risk through information literacy, promoting better human-computer collaboration and social development.

**LITERATURE REVIEW**

**Generative AI empowering libraries:** Nowadays, generative AI is the most promising development direction of AI, scholars in the field of libraries at home and abroad are also competing to explore the impact of generative AI on libraries and the prospects for the application of generative AI in the field of libraries. It has a unique function in typical library situations like smart Q&A and reference counselling, Scholars in the library field have also keenly seized this opportunity with a series of discussions around generative AI. Fu Rongxin et al. (Fu et al., 2023) tried the application effect of three generative AI products, ChatGPT, ERNIE Bot, and Bard, in library scenarios and found that all of them can play an advantageous role in library-related scenarios, the difference lies in the different applicability of different scenarios. Wang Nuo et al. (Wang et al., 2023) proposed the potential application scenarios of AIGC in the field of graphical information. They argued that AIGC could build the virtual space of libraries, re-engineer internal organizational structure, reshape external interactions, explore the value of collection resources, and expand public service functions. Li Guangzhu (Li et al., 2023) describes the dividends brought by ChatGPT to university libraries in terms of promoting the value-added efficiency of library business, boosting the readers' service work to increase wisdom and intelligence, and innovating the quality education mode of libraries, and analyses the impact of ChatGPT on the future development of libraries. Das and Islam (Das et al., 2021) conducted a systematic literature review of 32 papers on the use of AI and ML in libraries, extracting nine themes of AI applications: collection building and management, circulation and user services, reference advisory services, library management, and library customization and retrieval. Cox et al. (Cox et al., 2023) discuss the potential impact of ChatGPT on academic libraries, arguing that ChatGPT can complement or even replace existing search methods to empower library reference counseling services and that libraries should evaluate, develop, and use it. Fernandez (Fernandez et al., 2023) provides a foundational understanding of the technology behind ChatGPT, summarizes the deep learning models that make ChatGPT work, explores some of the potential controversies, disinformation, etc., surrounding this technology, and analyses that library professionals should continue to be informed about generative AI technology concepts to address future changes.

**AI hallucination:** AI hallucinations are inaccurate, unrealistic, surreal, or even completely fictional images, text, sound, or video content generated by AI systems, especially large-scale models based on deep learning, in the absence of real data or valid constraints, through over-generalization, irrational extrapolation, or creative interpretation of their training data (Hatem et al., 2023).When the truth is revealed, generative AI that spreads misleading or erroneous information on which users could base their judgments may cause cognitive dissonance, unease, and bewilderment. This psychological conflict caused by misinformation increases significantly, especially during critical decision-making. Research has shown that information-needs mismatches can trigger similar cognitive and behavioral problems in question-and-answer communities. Therefore, ensuring the quality of information provided by AI is critical to maintaining user trust and reducing cognitive dissonance. Zhou Tao et al. (Zhou et al., 2024) studied the intermittent discontinued behavior of generative AI users based on C-A-C. Data analysis using SEM and fsQCA concluded that privacy concerns and information illusions influence cognitive dissonance, which leads to intermittent discontinued behaviors, suggesting the need to improve the system’s level of functionality such as intelligence, anthropomorphism and personalization. Chen Wanqiu et al. (Chen et al., 2024) provide an in-depth discussion on the phenomenon of "knowledge hallucination" and its associated risks induced by generative AI, such as ChatGPT. They analyze the types, characteristics, and causes of this phenomenon, highlighting the potential impact of hallucination risks. Furthermore, they propose strategies for risk management from three dimensions: technical, ethical, and social. Huang Lei et al. (Huang et al., 2023) conducted a comprehensive examination of hallucinations in Large Language Models (LLMs), addressing their rationale, classification, challenges, and unresolved issues. They present a new taxonomy of hallucinations, provide an in-depth analysis of contributing factors, and outline detection methods and benchmarks. Additionally, they discuss representative methods for mitigating hallucinations, identify current challenges, and pose open questions for future research to advance the reliability and safety of LLMs.

**Information literacy:** with the rapid advancement of information technology, information literacy education has become a core priority in the library field, and the rise of artificial intelligence technology has led to the emergence of AI literacy. As an extension of information literacy and digital literacy, the term “AI literacy” was coined as early as 1972 but has not received much attention until recent years. Scholars such as Cetindamar D. (Cetindamar et al., 2022) have defined AI literacy as a core competency encompassing a wide range of technologies, work processes, human-computer interactions, and learning, closely related to data literacy. In contrast, Long D. et al. (Long et al., 2020) emphasize that AI literacy involves the ability to critically evaluate AI technologies, effectively communicate and collaborate, and apply AI tools across various scenarios, all of which are grounded in digital literacy. In its 2020 statement, the International Federation of Library Associations (IFLA) advocated for libraries to use AI and machine learning to innovate their services and clarified the responsibility of libraries in AI literacy education. Several scholars, such as Lei Xiaoyan et al. (Lei et al., 2023), based on the technical level combined with the technical deconstruction and application value of super-large-scale cross-modal intelligent algorithmic models, depict the modular composition and risk countermeasures of AI-generated content embedded in digital literacy education under the big model, which will provide strong support for the construction of a universal and inclusive digital literacy education mechanism. Wu Dan et al. (Wu et al., 2022) traced back algorithmic literacy from related concepts such as information literacy, analyzed its connotative features, and constructed a competency framework for algorithmic literacy. Wen Fangfang (Wen et al., 2023) conducted a comparative study of five college libraries at home and abroad regarding information literacy cultivation objectives, teaching content, and teaching methods. On this basis, suggestions are made to expand the breadth and depth of the content of the general courses under the framework of information literacy competence, to promote the content of library information literacy general courses to keep pace with the times and to innovate and reform the teaching methods of information literacy general courses. Min Kexiang et al. (Min et al., 2024) analyzed the AI literacy education activities in Shanghai libraries through field research and network research, summarized the experiences and shortcomings, thought about the development of AI literacy education in public libraries, and put forward the corresponding strategies. AI tools such as ChatGPT have profoundly impacted areas such as libraries, and the authenticity and logic of their output have become a focus of attention. Clarifying how to screen and analyze the resultant responses of ChatGPT output, exploring the risks that may exist in its application, such as the existence of hallucinations and false information in the logic of language generation, and how to propose strategies to avoid and resolve these risks from the perspective of information literacy, will help those in the field to better cope with the opportunities and challenges under the impact of the new-generation of AI technology.

**INFORMATION ILLUSION IN GENERATIVE AI FOR LIBRARY APPLICATIONS**

**TYPES OF AI HALLUCINATIONS**

Large Language Models (LLMs), such as ChatGPT and Gemini, are AI systems that can read and generate natural human language, and their working principle is completely different from that of the human brain, which is only based on the pattern, structure, and probability of language. When interacting with such generative AI, users may get false content that seems true and reasonable but actually deviates from the intent of the question itself or even contradicts real-world knowledge, which is the phenomenon of “information illusion”.

According to the literature and related sources, it is confirmed that AI hallucinations have become universal and cover two main categories, namely Factuality hallucination and Faithfulness hallucination. Factuality hallucination refers to the inconsistency between the content generated by AI models and verifiable real-world facts, which can be further subdivided into factual inconsistency (i.e., contradicting real-world information) and factual fabrication (i.e., generating content that is completely unverifiable based on real-world information). The Faithfulness hallucination, on the other hand, focuses on the inconsistency between the model output and the user’s instructions or context, which can be further subdivided into instruction inconsistency, context inconsistency, and logical inconsistency (Huang et al., 2023).Further analyzed in the context of the Large Language Model, AI hallucinations can be divided into three categories: input-conflicting hallucination, context-conflicting hallucination and fact-conflicting hallucination. Input-conflicting hallucination refers to AI-generated content that does not match the user’s input references and usually occurs when the LLM does not accurately understand or follow the user’s input instructions. Context-conflicting hallucination is when AI-generated content contradicts previously generated information and may fail to maintain internal logical consistency or deviate from contextual understanding during multiple rounds of dialogue or long text generation. Fact-conflicting hallucination refers to the possibility that LLM may provide inaccurate, erroneous, or fictitious information when answering questions or generating text. Since the false information supplied by Fact-conflicting hallucinations can greatly affect the use of generative AI, current research focuses mainly on Fact-conflicting hallucinations (Qian et al., 2023).

**AI HALLUCINATIONS IN LIBRARIES**

The categorization of AI hallucinations varies across different research efforts. In libraries, generative AI can be used to produce a wide range of content, including text, images, audio, and video. It can also create digital products, virtual scenes, and digital characters, as well as generate and transform multimodal data. For example, Zhejiang Provincial Library has designed an intelligent customer service platform for AIGC scenarios, covering online workstations (supporting real-time consultation by librarians) and smart robots (with semantic understanding, natural language processing, multi-round dialogue, and knowledge Q&A capabilities). The Chat Library AI intelligent service platform, developed by the University of Science and Technology of China (USTC) library, comprises seven functional modules: Chat Library, Chat Language, Chat Fellows, Chat Documents, Chat Data, Chat Research, Chat Open, and Chat Academic. Users can use generative AI for intelligent consultation, translation, writing, and many other aspects to achieve efficiency and intelligence. Facing the application scenarios of generative AI-enabled libraries and extending the concept of AI hallucination to problems or phenomena related to AI technology that may be encountered in libraries, the four key application scenarios of text generation, image recognition, voice interaction, and recommended system can be analyzed in depth:

**(1) Text generation hallucination** occurs when generative AI processes linguistic information and, due to algorithmic limitations or data bias, produces textual content that is erroneous, confusing, or stylistically inappropriate. For example, it might incorrectly identify the author of a particular book or even create entirely fictional books, authors, or references. Additionally, AI may experience semantic confusion, substituting similar but distinct concepts or words for one another when understanding and generating text.

(2) **Image recognition hallucination** occurs when AI systems recognize book covers, author photos, or library layouts, and due to factors such as image quality, lighting conditions, and algorithmic limitations, produce misrecognitions. This can result in incorrectly classifying one image as another or, even worse, generating images of book covers or library scenes that appear real but do not exist.

(3) **Voice interaction hallucination** refers to the fact that in library AI voice services, due to the limitations of speech recognition and semantic understanding technology, comprehension bias occurs, and the AI is unable to accurately understand and respond to the user's voice commands, misinterpreting the user’s question as something completely different; when answering, it may also give an answer that is completely irrelevant to the user’s question, reducing the efficiency and accuracy of the interaction.

(4) **Recommended system hallucination** primarily manifests as inaccurate and biased recommendations. Based on flawed data or algorithms, AI may suggest books to users that are entirely irrelevant to their interests or needs. Additionally, due to biases inherent in the training data, the AI's recommendations may unfairly or discriminatively reflect these biases, disregarding or undervaluing the needs and preferences of certain user groups.

Among them, Recommended system hallucination has a greater impact on libraries. As a core function of library services, the recommendation system significantly impacts the efficiency and satisfaction of readers' information access. An effective recommendation system accurately aligns with readers' interests and needs, providing personalized reading suggestions. This not only enhances the reading experience but also fosters greater reader engagement and loyalty. The mismatch between the recommended content and readers’ interests, the existence of biased or discriminatory recommendations, etc., will directly lead to readers’ dissatisfaction and disappointment with library services, affecting the overall image and attractiveness of libraries. Detailed categories and examples of AI hallucinations in the library are shown in Table 1. Content labeled in blue represents hallucination output content, while content labeled in red indicates user instructions or provides context that contradicts the AI hallucination.

Table 1. Categories and examples of AI hallucinations in libraries.

|  |  |  |  |
| --- | --- | --- | --- |
| **Types of AI hallucinations** | **AI hallucination subtypes** | **Descriptive** | **Example** |
| Text Generation Hallucinations | Factual errors | AI provides incorrect factual information when  generating book descriptions, summaries, or recommendations. | AI-generated book summaries **incorrectly**  **state that a book was written by another author.** |
| Fiction | AI creates books, authors, or quotes that don't exist at all. | **AI recommends a fictionalized**, classic work of literature that doesn't actually exist. |
| Semantic obfuscation | AI confuses similar concepts or words when understanding and generating text. | **AI confuses the concepts** of “librarianship” and “library automation”. |
| Image Recognition Hallucinations | Misidentified images | AI produces errors when recognizing book covers, author photos, or library layouts. | **AI misidentifies a book** cover about animals as another book entirely. |
| Fictitious image | AI generates images of book covers or library scenes that don't exist. | AI generates what appears to be a real cover image **for a fictional book.** |
| Voice Interaction Hallucinations | Misinterpretation of instructions | The AI misinterprets the user's query or request in the voice interaction. | The **user asks** “Where are the latest history books?”**.** which the **AI misinterprets as** “the latest sci-fi book”. |
| Irrelevant answer | The AI gives answers that are not relevant to the user's question. | The user asks for library opening hours and the AI answers with information about the rules for checking out books. |
| Recommended System Hallucinations | Inaccurate recommendations | AI recommendation systems provide irrelevant book recommendations based on inaccurate data or algorithms. | The AI recommends a wide range of historical books to users who like science fiction. |
| Bias Recommendations | AI recommender systems are subject to training data bias, leading to unfair or discriminatory recommendation results. | AI is more inclined to recommend specific types of books to certain groups of users **while ignoring others.** |

**ANALYSIS OF THE CAUSES OF AI HALLUCINATIONS AND RISK IDENTIFICATION**

**ANALYSIS OF THE CAUSES OF AI HALLUCINATION**

**Data Deficiencies:** Analyzing the technical mechanisms, library AI systems rely on extensive pretrained data to construct their knowledge base and decision-making models. However, databases originating from diverse sources, such as the Internet and library repositories, inevitably contain erroneous, outdated, incomplete, or biased information. The quality of this data can directly impact the accuracy and reliability of AI systems, potentially leading to "hallucinations" when processing user queries, resulting in the provision of false or misleading information (Bauer et al., 2023). Generative AI may lack the ability to make informed decisions consistent with human values and preferences, constituting unforeseen errors and behaviors. For example, information about certain groups in training datasets is oversimplified or negatively portrayed, and AI-generated content may unconsciously reflect these biases, resulting information illusion.

**Algorithmic limitations:** AI systems often struggle to capture the nuances of knowledge boundaries when processing complex and multidimensional information due to algorithmic limitations or incomplete data. When dealing with texts or contexts requiring in-depth understanding and analysis, generative AI may inaccurately capture the nuances and complexities of the information, leading to an imprecise yet overconfident sense of factual knowledge boundaries. This overconfidence can mislead generative AI models to fabricate unfounded and deterministic answers. AI models often struggle to accurately capture the deeper meanings and contextual information of complex language. This comprehension bias can lead to model-generated content that is illogical or misaligned with user intent.

**The opacity of AI decision-making:** Decisions made by AI systems typically arise from machine learning algorithms that recognize patterns within datasets, rather than from direct human programming. This pattern recognition mechanism often makes the logical processes of AI opaque and challenging for humans to understand, raising issues of interpretability and transparency in algorithmic decision-making. The opacity of AI systems can lead to ambiguous attribution of responsibility in cases of misleading or incorrect information. It becomes unclear whether accountability should rest with the algorithmic designers, the data providers, or the AI systems themselves. This lack of a clear accountability mechanism further exacerbates the risk of AI hallucinations.

**User over-reliance on AI-generated content:** Generative AI, with its advanced natural language processing capabilities and efficient information integration, is increasingly becoming a key method for acquiring knowledge. However, the convenience and efficiency of generative AI may lead users to become overly dependent on AI-generated content, potentially undermining the importance of traditional information retrieval methods and critical thinking. This over-reliance can diminish users' initiative and critical thinking skills in processing information. Users might neglect to actively screen, analyze, and evaluate the authenticity and value of information, instead accepting AI recommendations and judgments uncritically. This dependency contributes to the exacerbation of AI hallucinations.

**RISK IDENTIFICATION OF AI HALLUCINATIONS**

**The Individual Level: Deepening Personal Cognitive and Decision-Making Bias.**

In exploring the problem of “AI hallucinations” in libraries applying generative AI, it is important to recognize the profound impact on individuals’ cognitive abilities, foundations of trust, and ethical attitudes.

Although users can achieve efficient information retrieval, reading recommendations, and answering questions when using generative AI-assisted library services, which greatly reduces the cost and facilitates knowledge searching, the content it generates lacks true comprehension and is merely a recombination of data. This apparent convenience hides risks, as users unconsciously accept false or misleading information during frequent use, leading to cognitive degradation and affecting the accuracy of daily decision-making.

False moral perceptions or biases resulting from AI hallucinations may obscure individuals' understanding of good and evil, challenging traditional moral concepts. Since AI lacks the capacity for autonomous moral and ethical judgment, the content it generates or recommends may include moral ambiguities or conflicts. This can lead to confusion in individuals' moral judgments and may even prompt some to engage in extreme or irresponsible behavior.

Users’ over-reliance on AI-generated content can undermine personal autonomy. Additionally, generative AI often delivers customized content based on users' interests and behaviors, enhancing the user experience but also exacerbating the phenomenon of information cocooning. Users are only exposed to information with similar views to their own, and the lack of collision and exchange of multiple perspectives will further solidify the original bias and cognitive framework, limiting the breadth and depth of thinking.

**The social dimension: knowledge dissemination and distortion of public discourse.**

With the widespread adoption of generative AI-assisted information services in libraries, the impact of AI hallucination is gradually expanding. Deviations or errors in AI comprehension and generation of natural language can result in misleading, biased, or even fictionalized information. These inaccuracies can adversely impact the construction of public discourse and the development of social cognition.

When library AI-generated content is widely disseminated and used as a source of knowledge, the potential hallucination phenomenon can spread like ripples throughout the information ecosystem. This will confuse the public’s understanding of the truth and exacerbate the fragmentation and polarization of social information. Rapidly disseminating false or biased information on social media, news platforms, and other channels can quickly build a seemingly reasonable but distorted public discourse system, which can mislead the public's perception and judgment of complex social issues.

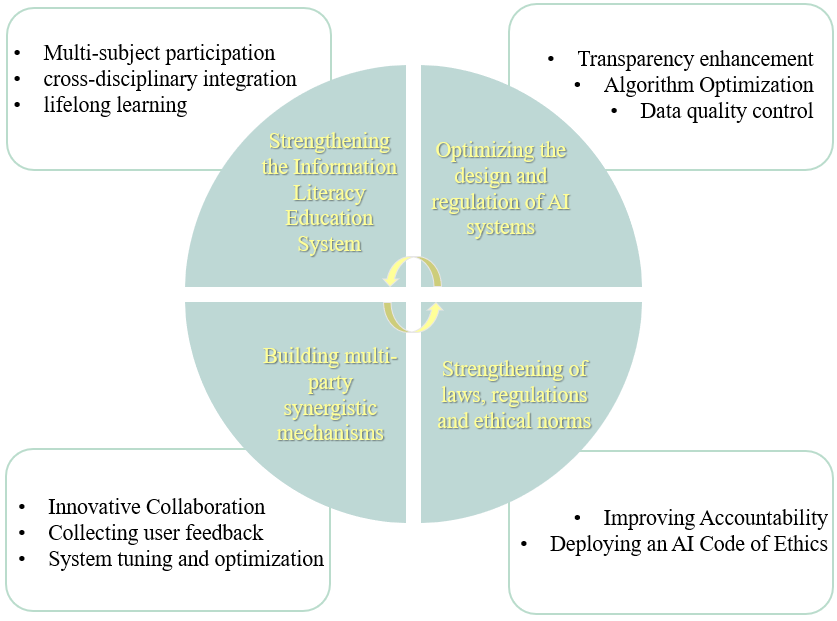
In addition, the hallucination phenomenon in AI services of libraries, as important institutions for knowledge dissemination, may also weaken the public's trust in authoritative information sources. Frequent occurrences of AI-generated misinformation can lead to widespread skepticism about the entire information ecosystem, causing traditional knowledge institutions, including libraries, to face a crisis of trust. This erosion of confidence can severely impede the effective dissemination of knowledge and hinder the healthy development of society.

Distortions in knowledge dissemination and public discourse can also amplify social inequalities and divisions. Fueled by information asymmetry and cognitive biases, these distortions can reinforce cognitive barriers and prejudices between different social groups, impeding cross-group communication and understanding. Specifically, individuals who lack information literacy or technological resources are more vulnerable to misinformation, which exacerbates the knowledge divide. Information illusion can also serve as a breeding ground for the dissemination of undesirable content, potentially manipulating public opinion and spreading false information. This poses a significant threat to social stability and public security. Therefore, from a social perspective, the risk of AI hallucinations in libraries extends beyond individual cognition and behavior to impact the stability of social structures and the protection of public interests.

**RISK AVOIDANCE STRATEGIES BASED ON INFORMATION LITERACY**

Generative AI, as a disruptive technology, is ushering in a transformative shift in information services, including libraries. Its possible risks in the application process, such as the existence of hallucination and false information in the logic of language generation, have already triggered a great uneasiness in the public’s heart.

From the perspective of information literacy, we propose a four-pronged strategy of strengthening the information literacy education system, optimizing the design and regulation of AI systems, constructing a multi-party collaborative mechanism, and strengthening the construction of laws, regulations, and ethical norms to avoid and resolve these risks, so that the public can better cope with the opportunities and challenges under the impact of the new-generation AI technology.

****

**Figure 1. Risk avoidance strategies under information literacy.**

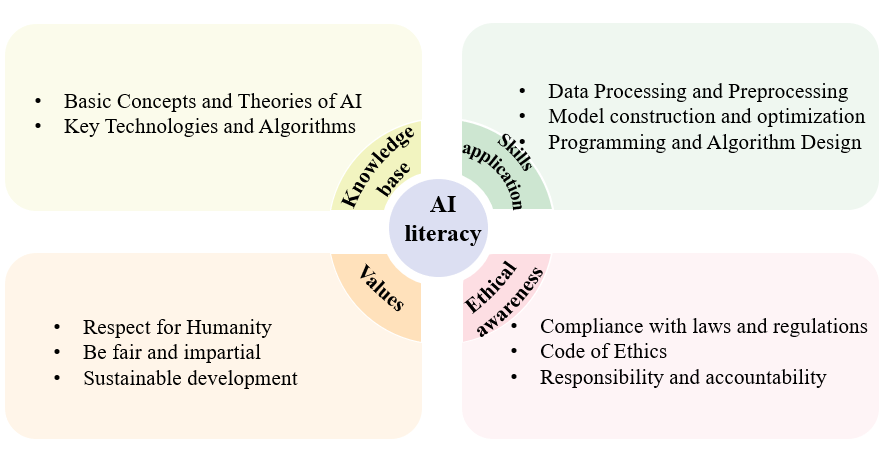
**STRENGTHENING THE INFORMATION LITERACY EDUCATION SYSTEM**

To effectively mitigate the risk of information illusion and enhance the quality of information services in the AI era, it is essential for libraries to strengthen their information literacy education system. The AI literacy framework is constructed based on the KSAVE model, initially used to assess 21st-century skills, and comprehensively covers the five key areas of knowledge, skills, attitudes, ethics, and values (Binkley et al., 2012). This paper summarizes the AI literacy framework into four dimensions: knowledge base, skills application, values, and ethical awareness. Based on this framework, an AI literacy education program is developed, which is committed to building an ecosystem involving librarians, educators, technology developers, policymakers, and the public to promote the deep integration of information literacy and AI literacy through interdisciplinary academic exchanges and collaborations.

Implementing comprehensive and tiered educational strategies is crucial for enhancing both theoretical understanding and practical familiarity with AI. Initially, it is imperative to focus on the training and education of librarians, who play a pivotal role in guiding users through AI-generated content. Librarians need to acquire a solid grasp of AI principles and fundamentals, as well as stay informed about key AI technologies. This will enable them to provide context and expert insights, effectively navigating the complexities of artificial intelligence and confidently mastering its applications.

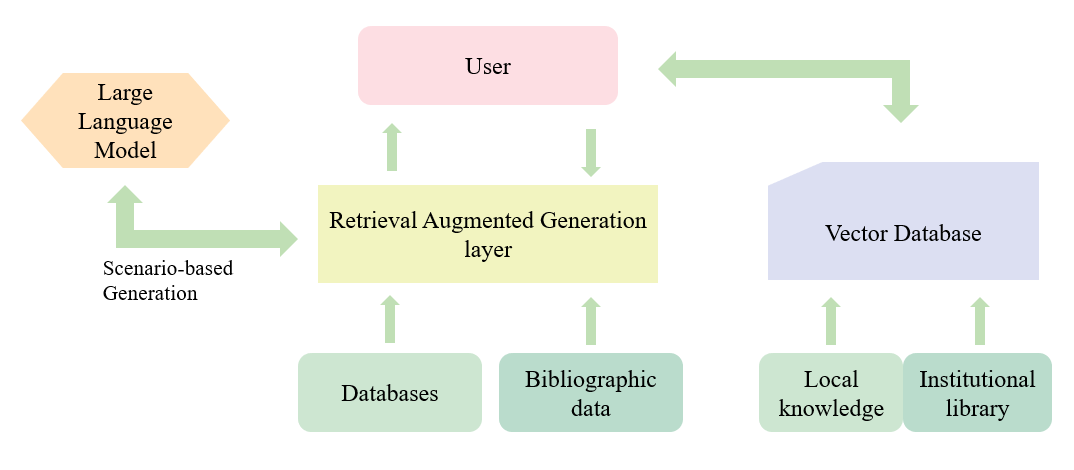
To enhance public awareness and education regarding AI-generated content, libraries should develop digital resources and intelligent learning platforms that offer professional guidance and support on AI information resources. Integrating AI technology with humanities, social sciences, and natural sciences to create an interdisciplinary knowledge system will aid the public in better understanding AI. For instance, Sichuan University has established a dedicated web page on generative AI, providing extensive information to improve teachers' and students' comprehension and application of the technology. Libraries can also integrate critical thinking and fact-checking skills into their information literacy courses, offer relevant lectures to cultivate essential assessment skills, and promote a culture of critical engagement and empowerment when interacting with AI-generated content.

Finally, establishing a lifelong learning system that supports ongoing education and self-improvement is essential for adapting to future changes in the information environment. Encouraging learners to stay current, update their knowledge base, and continually enhance their skills represents a sustainable approach to mitigating the risks associated with information illusion in AI language generation.

**Figure 2 .AI Literacy Framework.**

**OPTIMIZING THE DESIGN AND REGULATION OF AI SYSTEMS**

In the application of generative AI models in libraries, the root causes of information hallucination often lie in the subtleties of data quality and algorithm design. One contributing factor is the vast scale of datasets. In the face of diverse and complex training data, libraries can leverage Retrieval-Augmented Generation (RAG) technology. This approach integrates external document corpora into large language models during text generation, enhancing the accuracy, timeliness, and factual grounding of the model's outputs. By employing RAG technology, libraries can effectively filter out low-quality data, thus reducing the risk of hallucinations caused by data biases. Additionally, constructing a diversified data source system and promoting data diversity can help AI systems achieve a more comprehensive understanding of the information landscape, thereby mitigating the limitations imposed by single-source data (Sinha et al., 2022).

****

**Figure 3. Steps of RAG to build library service.**

The algorithms must be continuously iterated and optimized since AI language generation models are susceptible to multiple factors, such as training data distribution, algorithmic architecture, and hyperparameter settings. This includes the introduction of advanced deep learning algorithms, such as the Transformer model and its variants, to enhance the naturalness and accuracy of Text generation; the implementation of regularization techniques to prevent over-fitting of the model and improve the generalization ability; and the use of methods such as adversarial training to improve the model’s resistance to noise and hallucination information. In addition, attention should also be paid to the interpretability of the algorithms by designing a transparent algorithmic structure that makes the decision-making process of the AI system traceable and verifiable and provides strong support for the identification and correction of information illusions (Ramesh et al., 2022).

Advocating for greater transparency in library generative AI systems means disclosing information about algorithmic logic, data usage, and decision-making criteria to users, administrators, and the broader community. Ensuring transparency in AI-generated content involves clarifying the reasoning behind the content and verifying its sources. This approach enhances public trust in AI technology and supports adherence to legal and ethical standards in its application. Transparency can be enhanced through several measures: establishing comprehensive instruction manuals for AI systems, providing visual representations of algorithmic decision-making processes, and conducting public education on AI technology. Additionally, fostering interdisciplinary collaboration by involving experts in law and ethics in the design and evaluation of AI systems ensures adherence to legal standards and ethical norms.

**CONSTRUCTING MULTI-PARTY SYNERGY MECHANISMS**

Algorithmic limitations are a significant factor contributing to information illusion. Libraries can address this issue by partnering with technology providers to jointly invest in the research, development, and innovation of AI technology. Such collaborations can facilitate in-depth investigations into the specific problems associated with AI hallucinations and lead to the formulation of more effective solutions. Through innovative collaborations, libraries can share technological resources with partners, including technology providers and academic institutions. These resources encompass advanced algorithmic models, high-quality datasets, and expert technical teams. Such sharing enables libraries to gain more comprehensive technical support, optimize the performance of AI language generation systems, and mitigate the risk of hallucinations arising from technical limitations. Organizations such as library associations and trade associations often collect and disseminate best practice examples both within and beyond their industry. By participating in these networks, they can learn from the experiences of others and avoid repeating similar mistakes.

As the primary users of AI language generation systems, individuals play an essential role in enhancing system performance and mitigating the risk of information illusion through their feedback. To establish an effective user feedback evaluation mechanism, libraries should prioritize a user-centered approach by creating robust feedback channels and evaluation systems. Through online surveys, user interviews, social media monitoring, and other methods, the library should widely collect users’ opinions and suggestions on AI language generation services to ensure the comprehensiveness and timeliness of the feedback information. Based on this feedback, libraries should employ advanced technologies such as big data analytics and text mining to conduct in-depth analysis. This analysis will help identify key issues and potential risks that users are concerned about. The results should be promptly communicated to the technical team and business departments to guide targeted improvements and optimizations. Additionally, libraries should keep users informed about how their feedback has influenced updates to AI language generation services, which can help pinpoint potential hallucination risk areas. Establishing a user feedback incentive mechanism is also crucial to encourage active participation, thereby enhancing user satisfaction and loyalty. This approach fosters a virtuous cycle, boosting user engagement and responsibility while contributing to the overall improvement and healthy development of the library's AI language generation system.

**STRENGTHENING LAWS, REGULATIONS, AND ETHICAL NORMS**

To address the challenge of knowledge hallucination caused by generative AI, libraries must develop a comprehensive ethical governance framework to ensure the technology's ethical and safe application. This entails a thorough analysis of generative AI's characteristics and its role in library services, the formulation of actionable ethical guidelines, and the establishment of clear ethical standards. Such measures are essential to safeguarding users' rights and interests while preventing the dissemination of misleading information and protecting public trust. Simultaneously, it is crucial to establish a hierarchical regulatory mechanism to manage various types of knowledge generation services with appropriate differentiation. Libraries should also engage in public awareness campaigns about knowledge hallucinations by organizing lectures, seminars, and online courses. These educational activities aim to disseminate knowledge on AI ethics and foster an ethical consensus among academics, policymakers, and the general public. Additionally, libraries should actively contribute to the development of ethical principles and policy recommendations, offering a scientific foundation for policy-making. Within institutions, libraries need to establish ethics committees or designated departments responsible for overseeing AI technology application. These bodies should assess risks, develop countermeasures, and collaborate with research institutes and enterprises to jointly formulate industry-wide ethical guidelines. Such collaboration will also enhance information sharing and ensure comprehensive ethical oversight in AI implementations.

To realize agile governance, libraries need to flexibly adjust their governance strategies, establish a rapid response mechanism to deal with potential risks, and adopt governance tools that combine prudence and flexibility to ensure the effectiveness and adaptability of governance measures. Facing the complexity and multi-subjectivity of the risk of knowledge hallucination, libraries should start from a collectivist perspective, strengthen synergistic cooperation with the government, scientific research institutes, industry associations, etc., establish a multi-subjective responsibility mechanism, build a human-machine consortium, and enhance the integration of resources to deal with the responsibility problem together (Hacker et al., 2020). By implementing these measures, libraries can effectively mitigate the negative effects of generative AI knowledge hallucinations, foster the responsible development and application of the technology, and establish a clear accountability system. This approach will ensure the rigorous enforcement and oversight of ethical standards, promoting a balanced and ethical integration of AI in library services.

**CONCLUSION**

The emergence and advancement of generative AI are poised to revolutionize the library service industry, shifting libraries from traditional roles to encompass knowledge and generative content services. As AI technology evolves, it introduces risks such as AI hallucinations, which can lead to the dissemination of incorrect information. Consequently, the significance of information literacy will become increasingly critical, underscoring the need for robust educational frameworks to navigate the evolving landscape of AI-driven library services. Therefore, it is essential to continually enhance information literacy and deepen the understanding of AI technology to safeguard library users' rights, uphold the integrity of library information, and foster effective human-machine collaboration. While the current study offers valuable insights, it is limited by its scope and calls for empirical research to validate its findings. Future research should focus on evaluating the effectiveness of the proposed strategies, exploring new methods for risk mitigation, and continually innovating to advance public information literacy. This will contribute to developing more comprehensive solutions and offer new perspectives for improving the synergy between humans and AI in library settings.

**REFERENCES**

Fu, R. X., & Yang, X. H. (2023). Analysis of AIGC Language Models and Research on Their Application Scenarios in University Libraries. Journal of Library and Information Sciences in Agriculture, 35(7), 27-38.

Wang, N., Bi, X. C., & Xu, X. (2023). Sharpening the Tool First: AIGC in the Metaverse and Its Opportunities for GLAM Applications. Library Tribune, 43(2), 117-124.

Li, G. Z. (2023). The Impact of ChatGPT on University Libraries: Operations, Services, and Development Paths. Library Theory and Practice, (6), 123-128.

Das, R. K., & Islam, M. S. U. (2021). Application of artificial intelligence and machine learning in libraries: a systematic review. arXiv preprint arXiv:2112.04573.

Cox, C., & Tzoc, E. (2023). ChatGPT: Implications for academic libraries. College & Research Libraries News, 84(3), 99.

Fernandez, P. (2023). “Through the looking glass: envisioning new library technologies” AI-text generators as explained by ChatGPT. Library hi tech news, 40(3), 11-14.

Hatem, R., Simmons, B., & Thornton, J. E. (2023). A call to address AI “hallucinations” and how healthcare professionals can mitigate their risks. Cureus, 15(9).

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence‐written research papers and the ethics of the large language models in scholarly publishing. Journal of the Association for Information Science and Technology, 74(5), 570-581.

Zhou, T., Zhang, C. L., & Deng, S. L. (2024). Research on Intermittent Discontinuance Behavior of Generative AI Users Based on C-A-C. Modern Information, 1-17. [Online]. Retrieved from http://kns.cnki.net/kcms/detail/22.1182.G3.20240710.1235.002.html [Accessed on July 20, 2024].

Chen, W. Q., & Luo, Y. R. (2024). The "Knowledge Illusion" of Generative Artificial Intelligence and Its Risk Governance Exploration. Journal of the Shanghai Institute of Socialism, (04), 38-51.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. arXiv preprint arXiv:2311.05232.

Cetindamar, D., Kitto, K., Wu, M., Zhang, Y., Abedin, B., & Knight, S. (2022). Explicating AI literacy of employees at digital workplaces. IEEE transactions on engineering management, 71, 810-823.

Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-16).

Lei, X., & Shao, B. (2023). Embedding artificial intelligence-generated content into digital literacy education under the large model framework. Modern Information, 43(06), 99-107.

Wu, D., & Liu, J. (2022). Algorithmic literacy in the era of artificial intelligence: An analysis of its connotation and the construction of a competency framework. Journal of Library Science in China, 48(06), 43-56.

Wen, F. F., & Liu, Y. (2023). Reflections on information literacy general education in Chinese university libraries: A comparative study based on 10 libraries in China and abroad. Library Theory and Practice, (01), 114-122.

Min, K. X., & Feng, Y. (2024). Empowering the future: Practice and reflections on artificial intelligence literacy education at the Shanghai Library. Journal of Shanxi Library Science, 1-14.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. arXiv preprint arXiv:2311.05232.

Qian, Y., & Mei, Y. (2023). From concept to practice: Exploring the application of generative artificial intelligence in smart libraries. Library Research and Work, 12, 27-34.

Bauer, K., von Zahn, M., & Hinz, O. (2023). Expl (AI) ned: The impact of explainable artificial intelligence on users’ information processing. Information systems research, 34(4), 1582-1602.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. Assessment and teaching of 21st century skills, 17-66.

Sinha, A. R., Goyal, N., Dhamnani, S., Asija, T., Dubey, R. K., Raja, M. V., & Theocharous, G. (2022). Personalized Detection of Cognitive Biases in Actions of Users from Their Logs: Anchoring and Recency Biases. arXiv preprint arXiv:2206.15129.

Ramesh, K., KhudaBukhsh, A. R., & Kumar, S. (2022, June). ‘Beach’to ‘Bitch’: Inadvertent Unsafe Transcription of Kids’ Content on YouTube. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 11, pp. 12108-12118).

Hacker, P., & Passoth, J. H. (2020, July). Varieties of AI explanations under the law. From the GDPR to the AIA, and beyond. In International workshop on extending explainable AI beyond deep models and classifiers (pp. 343-373). Cham: Springer International Publishing.