A Study on Risk Identification and Credibility Enhancement Pathways for Artificial Intelligence Literature Retrieval Tools

Lele Zhong
Library, Hangzhou Dianzi University, Hangzhou, Zhejiang, China

Abstract: The iterative progress of generative artificial intelligence technology is driving the development of academic literature retrieval tools towards intelligence, natural-language question answering, and deep knowledge integration, thus improving the efficiency of information acquisition for researchers. At the same time, there are also many credibility problems due to altered content and opaque algorithms. Given a clear definition of the connotations and scopes of AI-powered literature retrieval tools, four types of risks can be systematically identified in this paper: authenticity of content, algorithmic mechanisms, academic ethics, and data security. It will also investigate the reasons for the above risks at the levels of technology, behaviour and institutions. Based on the above analysis, a four-dimensional credibility evaluation system of content authenticity, algorithm openness, ethical compliance and user suitability has been established. Additionally, it proposes a combined method to strengthen the credibility of technology optimisation, platform governance, user literacy and institutional regulations. The research hopes to offer some theoretical support for the healthy development of AI-based literature retrieval tools and provide a practical reference for researchers ' responsible use of intelligent tools and academic publishers ' optimization of governance systems.

Keywords: Artificial Intelligence; Literature Retrieval; Credibility

1. Introduction

World-class scientific results continue to accumulate rapidly. According to the data in the Scopus database, the number of new academic papers published each year has exceeded 5 million. Traditional literature search tools based on keyword matching have problems such as low retrieval efficiency, weak knowledge integration and difficulty in discovering implicit connections; as a result, they are time-consuming for researchers to use in literature screening and review synthesis. The development of large language models and retrieval-augmented generation technology has advanced the field of intelligent academic literature research. At present, some specialised academic AI search tools have been released, such as Scopus AI, Web of Science AI, Consensus and Elicit. Databases in China, such as CNKI and Wanfang Data, have added intelligent question-and-answer and AI-based literature review functions, and general-purpose large models have also begun to be used by many researchers for the first-round screening of literature.

AI literature retrieval tools can quickly provide all-around conclusions through natural language interaction, automatically organise research paths, and recommend relevant papers in bulk; thus, they have been regarded as indispensable tools for improving the efficiency of research by lowering the barrier to accessing academic information. However, with the expansion of their applications, serious problems of credibility have gradually emerged, such as false citations, misrepresentation of views, algorithmic bias, and copyright infringement. Some researchers have made citation errors in their academic papers due to excessive reliance on AI-generated content and thus drawn attention. Academically, the community is aware of the convenience of AI retrieval tools but also uncertain about their accuracy; thus, at present, the main obstacle to their widespread use in research has been the problem of reliability.

Given the above, three urgent problems need to be addressed: First, the risks of AI literature retrieval tools are multifaceted, but there is no systematic framework for risk identification at present; therefore, technical, ethical and institutional risks cannot be fully covered. Second, neither academia nor industry has established unified standards for the credibility

Copyright @ STEMM Institute Press

assessment of these tools, and thus users are unable to compare the reliability of different tools. Third, the existing improvement measures are fragmented, either being technical optimisations or user education, and a systematic governance system has not been built. Therefore, this paper will address the problems of risk and credibility in AI literature retrieval tools. Normative analysis is used to build a risk framework and evaluation system, propose an all-round improvement path, and provide theoretical support for the healthy development of the academic information ecosystem.

Currently, most research has been conducted in three areas: the technical feasibility of AI literature retrieval tools and their applications [1,2]. These tools can be used for news text mining to improve the early warning of credit risk [3] and provide data retrieval support for local legislation [4]. Second, studies on information distortion risks associated with AI literature retrieval tools [5-7], including algorithmic bias, algorithmic opacity, data bias, and the need for further verification of information reliability and authenticity [8]. Third, research on the risk governance of AI literature retrieval tools [9, 10], focusing on data governance, quality control, privacy protection and security measures [11].

Although previous studies have provided some support for this research, three main problems have not been solved. Most of the current studies focus on the application value and neglect risk governance. Most of the existing research on the advantages and application scenarios of AI retrieval tools has not addressed the issue of risk, and systematic studies on risk identification are lacking. Secondly, the content of the current studies is not comprehensive and fails to build a general system. Most studies have addressed hallucinations and copyright problems separately, failing to integrate content, algorithms, ethics and data risks in a single system, and thus have not explored how risk generation mechanisms work. Thirdly, in terms of outcomes, the current governance proposals lack an organised path. Most of the recommendations are either isolated technical improvements or normative demands, and there is no all-encompassing system to boost credibility for technology, platforms, users and institutions. To address these gaps, this paper follows a research logic of "form analysis—risk identification—mechanism exploration—system construction—pathway proposal", first, defining core concepts of AI literature retrieval tools and clarifying their technical forms and operational logics; second, identifying risk types from multiple dimensions and revealing the underlying mechanisms of risk generation at the technological, actor, and institutional levels; third, constructing a four-dimensional credibility evaluation framework; and finally, proposing a systematic pathway for improving credibility.

The main innovations of this paper are as follows: First, a multi-dimensional risk identification system has been constructed to address the problem of hallucination from four aspects rather than just one, and the basic generation mechanisms of these risks have been systematically investigated. Second, an all-encompassing solution of "evaluation system + improvement pathway" has been proposed to design an integrated enhancement strategy based on the results of the evaluation system, providing systematic references for the development of tools, platform governance, and industry regulations.

2. Technical Forms and Operational Logic of Artificial Intelligence Literature Retrieval Tools

2.1 Definition of Core Concepts

An artificial intelligence literature retrieval tool is an intelligent academic information system based on AI technology, including large language models, vector databases and knowledge graphs. It can understand the user's natural language query to perform semantic-level document matching, full-text content extraction, knowledge integration and summary generation. The new form is to generate answers directly rather than just retrieving related documents; thus, an organized response to a question can be created by combining various pieces of information from the literature and citing the sources.

Compared with the old keyword search tools, the four basic differences in the new AI retrieval system are: first, it's not based on character matching and Boolean logic; instead, it understands meaning and connects knowledge through AI; second, it outputs all-encompassing, generated knowledge rather than a list of individual documents; third, it can handle multi-turn conversations and follow-up questions in an AI system, unlike the one-way search of the

 Copyright @ STEMM Institute Press

past; fourth, it has many advanced functions for processing knowledge, such as summarization, comparison and inference, etc. It should be noted that the AI literature retrieval tools mentioned here include not only AI modules integrated into traditional academic databases such as CNKI and Web of Science, but also standalone AI academic search platforms like Consensus and Elicit, as well as general large language models that can perform literature retrieval and review generation. At the centre are AI tools for searching academic papers and organising research materials.

2.2 Mainstream Tool Types and Application Scenarios

Based on their functions and strengths, the present AI-powered literature search tools can be divided into four types. The first category consists of intelligent question-answering retrieval tools. At present, the most popular type allows users to pose research questions in natural language, and then the tool provides a summary answer based on existing literature along with citations to the relevant sources. Typical examples are Consensus, ScopusAI and CNKI's AI Q&A. These tools are mainly used for the background research and quick verification of specific questions, and to collect the main ideas from many papers. The second type is the literature review and thematic mapping tool. These tools help to organise the development path of a research problem, automatically find stages in field evolution, prominent research schools, leading scholars and publications, new trends, and research deficiencies, and thus generate an organised literature review framework. Litmaps and Research Rabbit's AI-powered review functions are typical examples; they can be used in the topic selection stage to conduct all-encompassing domain analysis and lay the groundwork for writing a review. The third category is Citation Recommendation and Expansion Tools. By using semantic similarity and knowledge association, these tools can find indirect connections to related but uncited works, thereby expanding the search scope for researchers and uncovering essential studies that have been overlooked in keyword searches. The fourth type is full-text analysis and cross-document question-answering functions. These tools can upload the full text of multiple papers, automatically extract key information such as research questions, methods, data and conclusions, and perform comparative queries on the documents. They are particularly suitable for in-depth analysis and data integration in the detailed reading stage.

2.3 Underlying Technical Architecture and Operating Mechanism

Most current mainstream AI literature retrieval tools are based on a technical structure that combines "large language models" with "retrieval-augmented generation (RAG)". The operating logic can be divided into three parts: First, document vectorisation and knowledge base construction. A tool that converts titles, abstracts and even the full content of a large amount of academic papers into high-dimensional vectors is stored in a vector database as a semantic search knowledge base. The quality of these vectors, the coverage of the document collection, and how often they are updated directly affect the completeness and accuracy of the retrieval results. Second, Semantic Retrieval and Context Recall. After the user submits the query, the system first converts the question into a vector and then uses an algorithm such as cosine similarity to find relevant document snippets in the vector database for retrieval as context. The reasons are that the recall rate and precision of retrieval are relatively low; therefore, the generated response will likely be inaccurate or incomplete. Thirdly, large model generation and result output. The retrieved document snippets and the user's query are jointly provided to a large language model, and then, based on this contextual information, a coherent and well-structured answer with proper citations is generated. Some tools can verify the facts of the generated content and format citations automatically.

Although RAG has reduced the problem of hallucination in pure generative large models, it is not completely free from such errors. On the other hand, there can be errors in semantic matching for vector retrieval, and thus irrelevant content may be returned. At the same time, the generation process of large models is probabilistic; thus, there may be over-extrapolation or incorrect integration of context, and even self-generated content when the context is lacking, resulting in distortion.

3. Risk Identification and Formation Mechanism of Artificial Intelligence Literature Retrieval Tools

Copyright @ STEMM Institute Press

3.1 Key Risk Identification

3.1.1 Risk of content authenticity

The primary and most widely reported risk of AI literature retrieval tools is content distortion, which affects the reliability of these tools. The first type of manifestation is bibliographic hallucinations and fabricated citations. This is the most severe type of distortion, and the tools generate completely fake references, such as fictional author names, paper titles, journal names, DOIs, and page numbers. Some of these false citations are formatted correctly and contain specific information, so they are very deceptive. Many empirical studies have shown that less than 40% of the academic citations generated by general large language models can be verified as genuine, and even specialised academic AI tools have problems with citation mismatch and misattribution. Second, Distorted Interpretations and Inaccurate Summaries. When producing content based on research, errors may occur; for example, some statements are taken out of context, overgeneralised, or reversed in cause and effect compared to the original work by scholars. For instance, they may be drawing unwarranted causal inferences from correlations under certain circumstances, ignoring the limitations of the study and future research directions, or selectively presenting evidence. The falsification is relatively minor compared to fabricated citations and cannot be identified without referring to the original documents. Thirdly, Coverage Bias and Information Omission. Due to limitations in training data and vector databases, most AI retrieval tools have shown a bias towards English-language publications, popular disciplines, and highly cited papers. As a result, many niche areas, non-English literature, research conducted by scholars in developing countries, grey literature and preprints have been neglected. Therefore, there will be systematic biases in the search results, and many important heterogeneous studies will be missed by the AI.

3.1.2 Algorithmic mechanism risks

Algorithmic risk at the level of the algorithm is hidden and structural, and its damage to academic research is serious. First, find black boxes and opaque ranking mechanisms. The traditional literature database has a relatively open sorting mechanism that can be customised according to different needs, such as relevance, citation frequency, publication time, etc. AI-powered retrieval tools are generally algorithmic black boxes for both content generation and result ranking; users cannot know how the output results are derived, so it is impossible to determine whether the tool gives more weight to authoritative or relevant sources, and potential commercial bias in the recommendations cannot be detected. Secondly, Information echo chambers and Homogenization Convergence. Most AI retrieval tools are based on personalised recommendations according to the user ' s previous search and reading history. Extended use restricts the scope of knowledge for students, limits exposure to cross-disciplinary studies and alternative viewpoints, and thus forms an academic information bubble. Therefore, there is a lack of academic criticism and innovation, a reinforcement of path dependence, and no pluralism in thought. Thirdly, delayed timeliness and omission of cutting-edge content. Large models are trained on datasets with a fixed cutoff date, and vector databases need to be updated periodically; as a result, most AI retrieval tools cannot access the latest published papers, working papers, and conference proceedings. For the new and developing fields, AI retrieval results may be up to 3-6 months out of date and therefore not suitable for the researchers' needs.

3.1.3 Academic ethics and normative risks

The spread of AI literature retrieval tools has disrupted the old norms and ethical systems in academia. First, there is a problem of academic dishonesty. Some scholars and students have been using AI to write the literature review and citations in their research without checking them. Therefore, there are many false citations and mistaken views that constitute covert academic dishonesty. In some cases, people have submitted entire AI-generated review papers and violated the rules against plagiarism and academic dishonesty. Second, knowledge attribution is unavailable. When AI tools combine information from many places to generate new ideas, it is often difficult to know where to find the original data for a particular claim. This will prevent the tracking of academic work and may also violate the author's right of authorship and intellectual property, thus causing an ethical problem. Thirdly, the academic Matthew effect is intensified. Algorithms are likely to promote well-cited and popular works and scholars, thus increasing the visibility of mainstream research and reducing

 Copyright @ STEMM Institute Press

the chances of discovery for niche studies, early-career researchers, and institutions at the margins. It will worsen the uneven distribution of academic resources and strengthen the existing academic hierarchy.

3.1.4 Data security and rights risks

Data-level risks directly affect the basic rights and academic freedom of researchers. First, privacy violations in user research: When researchers use AI retrieval tools, sensitive information such as the research questions entered, unpublished manuscripts uploaded, and ongoing project ideas may be collected by the tool provider for model training or commercial analysis, leading to risks of research idea theft and project direction disclosure. The risk of such a project is relatively high; thus, it is not recommended for now. Second, copyright compliance risks: Some AI retrieval tools, without permission from copyright holders, reproduce, parse and vectorise the full texts of copyrighted academic works for commercial question-and-answering services, thereby exceeding the scope of fair use for traditional work retrieval. Recently, many copyrights infringement cases have arisen as a result, and the development of the industry has been affected.

3.2 The Internal Mechanism of Risk Formation

3.2.1 Technical origins: The probabilistic nature of large model generation

At the root of the credibility risks for AI retrieval tools are technical attributes of large language models. At the bottom, these models are based on statistical probability learning of large amounts of text data and generate content by predicting the next character. They do not have genuine knowledge of facts or the ability to think logically; therefore, “hallucinations” are unavoidable. Although the addition of external knowledge bases in RAG architectures has reduced hallucinations to some extent, problems of content distortion still exist due to inaccuracies in vector search and context window limitations of the model. Furthermore, the academic literature is highly specialised, rigorous and context-dependent; thus, even minor changes in terminology or conditional clauses can lead to different conclusions and increase the risk of distortion in AI-generated content.

3.2.2 Primary driving force: Behavioural biases of the supply and demand sides

Behavioral biases at both the supply and demand sides have jointly increased the risk of contagion. Some platforms are driven by commercial interests on the supply side and, in pursuit of a good user experience and smooth interaction, have reduced the strictness of fact-checking for complete and fluent answers. Some tools do not inform users about their own deficiencies or are even intentionally presented as “all-knowing and precise”, causing people to expect too much from them. On the demand side, some researchers have shown cognitive biases by taking the AI ’ s output as the final answer without checking or expert judgment; others have used AI tools out of opportunism to avoid reading research papers and passively accepted incorrect information, thus becoming important links in the risk chain.

3.2.3 Institutional incentives: Lagging supply of governance rules

The absence of institutional rules and weak regulation are the main reasons for the failure to control risk. At present, there are no globally recognized industry standards specifically for AI-powered academic search tools, and no unified requirements or certification systems have been established for the basic indicators of data source compliance, citation accuracy, and algorithm transparency. Guidelines for the use of AI in academic publishing are still in the exploration stage; most journals require only a general disclosure of AI use and have not issued specific rules on how to use AI search tools or set criteria for identifying and penalizing false citations. At present, the regulations for generative AI are relatively general; specific governance measures for the particular features of academic information have not been introduced, and a cooperative governance system involving all parties has not yet been established.

4. Constructing a Credibility Evaluation System for Artificial Intelligence Literature Retrieval Tools

4.1 The Connotation and Evaluation Principles of Credibility

The reliability of an AI literature retrieval tool refers to the truthfulness and accuracy of its output and conclusions, the transparency and explainability of the operating process, and whether the service procedure meets standards

and responsibilities; thus, it can gain the trust of the academic community. At its foundation is the idea of "traceability, verifiability, explainability and accountability"; all generated content must have a reliable origin, the main claims can be verified by the original text, the operating logic is understandable to users, and clear responsibility is assigned in case of problems.

The four basic features of the construction of the credibility assessment system are as follows: First, scientific basis; indicators need to be in line with the technical traits and risk factors of AI retrieval tools and accurately reflect their level of reliability. Second, all-encompassing nature; cover all areas such as content, algorithms, ethics and users, and form a complete evaluation circuit. Third, convenience of application; the indicators should be relatively easy to quantify and observe to avoid excessive use of abstract qualitative descriptions. Fourth, flexibility: given the rapid development of technology, the indicators and their weights can be updated in the future.

4.2 Four-Dimensional Credibility Evaluation Index System

According to the above risk factors, this paper builds a four-dimensional credibility assessment system with four main indices and 16 sub-indices, as shown in Table 1. The three main rules for weight allocation are as follows: First, according to the risk-oriented principle, indicators with more serious consequences and a greater impact on credibility should be given higher weights. Second, in line with the coreness principle, core indicators that directly affect the authenticity of the content should be assigned a higher weight than process- or support-related indicators. Finally, under the baseline priority principle, compliance and legality baseline indicators are given higher weights than those related to user experience and optimisation.

Table 1. Four-Dimensional Credibility Evaluation Index System

Primary Indicator	Secondary indicators	Indicator Definition
Content fidelity	Accuracy of citations	The proportion of verifiable citation entries to the total number of citations
	Original fidelity	Output the degree of consistency between the viewpoint and the original meaning, verified through manual sampling
	Literature coverage	The proportion of core literature in the field covered by the search results
	Timeliness Match Rate	The proportion of the most recent literature within the past three years in results
Algorithmic Transparency	Explainability of sorting logic	Whether to disclose the core weighting rules for result generation and ranking
	Data source transparency	Is the source scope, subject coverage, and update cycle of the literature database clearly specified?
	Content traceability	Does it support clicking on a citation to directly jump to the corresponding paragraph in the original text?
	Degree of bias disclosure	Whether to proactively inform users about potential language or subject biases
Ethical Compliance	Copyright Compliance	Have you obtained legal authorization for the use of the document content?
	Privacy protection level	The scope of user data collection, storage duration, and the level of regulation regarding usage rules
	Level of bias control	The extent of inclusiveness and coverage for niche disciplines and non-mainstream research
	Sufficiency of Risk Disclosure	Does it significantly indicate the limitations of AI output and the obligation for human verification?
User Adaptability	Error Correction Feedback Mechanism	Does it support users submitting error feedback and enabling rapid iteration for fixes?
	Manual Review Channel	Is a value-added service with professional manual verification provided?
	Scene Adaptability	Does it support personalized settings for different disciplines and search scenarios?
	Convenience of verification	User's ease of verifying the original text

4.3 Explanation of the Evaluation Implementation Method

The evaluation system is composed of both qualitative and quantitative measures; first, according to the Analytic Hierarchy Process (AHP), experts in library and information science, AI technology and academic ethics are consulted to set the weights of indicators at all levels, and then, second, these indicators are scored through technical tests, manual checks and platform surveys; finally, the overall credibility score is obtained by means of the fuzzy comprehensive evaluation method, and the results are categorized into four grades: excellent, good, qualified and unqualified.

 Copyright @ STEMM Institute Press

The three kinds of applications for this system are as follows: first, tool developers use it to conduct self-assessments and iterative optimizations to address specific deficiencies; second, third-party organizations carry out industry evaluations to guide researchers in selecting tools; and third, universities and research institutions conduct tool eligibility assessments to screen for compliant and reliable tools for faculty and students. According to the progress of changes in technology and regulations, assess dynamically and continue to supervise regularly through evaluation activities.

5. Pathways to Enhancing the Credibility of Artificial Intelligence Literature Retrieval Tools

5.1 Technical Optimisation: Strengthening Traceability Verification and Data Quality

Technology provides the foundation for enhancing credibility and needs to address the root causes of distortion. First, fully implement a "sentence-level traceability" technical norm. Strengthen the fine-grained traceability of RAG architectures to link each generated sentence back to specific paragraphs and page numbers in the original documents, and avoid generating content without a verifiable source. Require that all citations be linked to genuine bibliographic metadata and direct links to the original works to prevent the generation of fake references. Second, build a multi-source cross-validation mechanism. Add a fact-checking module to cross-reference the main conclusions and required data from multiple sources, classify and label the content according to the level of confidence, automatically issue a red alert for suspicious content, and apply multiple verification rules to high-risk citation generation stages to ensure that every cited item exactly matches the original text. Thirdly, enhance the quality and scope of knowledge bases. Increase the inclusion of literature in less commonly used languages, niche disciplines, preprints and grey literature to reduce systematic coverage bias; establish a real-time update mechanism for the document library to shorten the data refresh cycle and improve coverage of cutting-edge research; and conduct domain-specific fine-tuning for different academic fields to enhance the accuracy of understanding specialized terminology and logical structure.

5.2 Platform Governance: Strengthening Full-Process Risk Management

The platform will take the lead in risk management and build an all-encompassing governance system. First, establish a system to label AI-generated content and assign confidence scores. All AI-generated works will be labelled with an "AI-generated" flag to distinguish them from human-created works. According to the extent of verification and the reliability of the source, the content is classified as high, medium or low confidence. Content from a single source that has not been verified should be labelled as low-confidence and thus fall under the scope of user notification requirements for risks. Second, set up a risk-tier control system and a manual-review mechanism. For high-risk scenarios, such as the use of academic theses, journal submission and grant application, add an extra pop-up warning to enhance verification. Establish a professional academic editing team to manually revise and correct frequently used core areas and high-scrutiny topics. Establish a high-speed error-reporting path to address issues in the displayed content promptly based on user reports. Thirdly, actively improve algorithm transparency. Disclose the technical architecture, data sources, update frequency, main algorithm logic and known defects of the tools to inform users of their rights to know. Stop using the "black-box marketing" method and, instead, objectively present the tool's strengths and weaknesses to avoid raising false hopes among users due to over-promotion.

5.3 User Empowerment: Strengthening Critical Use Literacy

The final line of defence for the credibility of the system is the user. therefore, users with low information literacy may fail to recognize and mitigate such risks. First, strengthen training on the proper use of AI academic tools. Universities and research institutions should add the principles, advantages, limitations and proper use of AI retrieval tools to their research integrity training programmes to provide specialised education for all faculty and students. Correct the misconception that "AI output is true" and promote the correct understanding that "AI is a tool to help; therefore, human supervision is required". Second, promote standardised use procedures. Promote a four-step method: "AI preliminary screening - manual

Copyright @ STEMM Institute Press

evaluation - original text verification - proper citation". Use AI only for the initial filtering of literature and a preliminary summary of ideas; verify the main references with the original materials and confirm all citations against their sources. Do not copy the review content or the list of references directly from the AI. Thirdly, foster the users ' critical thinking. Help the researchers strengthen their own evaluation of information, be cautious about accepting conclusions generated by AI, actively seek out various perspectives and conflicting studies, and avoid the echo chamber effect of algorithms. Promptly report any errors in the tool and participate in the iterative improvement process to form a positive feedback loop of user oversight.

5.4 Institutional Framework: Building a Multi-actor Collaborative Governance System

The first type of trust-building is provided by institutional norms; thus, many parties need to cooperate in governance. First, build an industry-standardised and trustworthy certification system. The Library and Information Science Association, in cooperation with academic publishing institutions, will issue a set of "Standards for Credibility of AI Academic Search Tools" to establish a minimum-requirement system for key indicators such as citation accuracy, algorithm transparency, and copyright compliance. A third-party credibility certification mechanism will be introduced to issue certified labels for compliant tools, and regular audits and dynamic exit mechanisms will be employed to promote continuous improvement in quality throughout the industry. Strengthen the regulation of academic publishing and research integrity. Academic journals will strengthen the rules for disclosing the use of AI tools and require authors to explicitly state whether AI search tools were employed and assume full responsibility for the accuracy of all content. Fabricated citations produced by AI will be considered a type of academic misconduct and will be dealt with more severely. Universities and research institutions need to strengthen the system of research integrity management and set rules for using AI tools. Thirdly, establish a multi-party cooperative governance system. The government will issue a specific document for the use of AI in educational data and outline the corresponding requirements. Strengthen self-regulation by the industry association and promote consensus in the sector. The platform company needs to focus on its main work and strengthen governance. Use AI tools responsibly and participate in the supervision of researchers. Together, these cooperative mechanisms of government supervision, industry self-regulation, platform responsibility and user participation will help to build a healthy academic information ecosystem.

6. Conclusion and Outlook

6.1 Primary Research Results

This paper conducts an organized study of the risks and credibility problems in artificial intelligence (AI) literature retrieval tools, and reaches the following main conclusions: First, AI-based literature retrieval tools are changing the model of academic information acquisition in an unprecedented way. To increase the efficiency of research, some credibility risks have arisen. In addition to the risk of content distortion, there are also four other types of risks: the authenticity of content, algorithms, academic ethics and data security; these risks may be explicit or implicit, and they also involve technology and institutions working together. Second, the credibility risk of AI retrieval tools is due to the combined effect of technical, actor and institutional factors. Probabilistic features of large models provide the technology foundation; behavioural biases of users and providers drive dissemination; delayed governance rules act as institutional triggers. The three are in combination and work together to create the current credibility problems. Thirdly, to improve the reliability of AI literature retrieval tools, some measures need to be taken. A four-dimensional evaluation system of content fidelity, algorithmic transparency, ethical compliance and user adaptability has been introduced to provide a practical foundation for credibility assessment. All-around improvements in technology, platform governance, user empowerment and institutional regulation have been introduced to establish a complete safeguard system for end-users.

6.2 Research Limitations and Future Directions

There are still some deficiencies in this study: it is mainly a normative analysis, the weights of

 Copyright @ STEMM Institute Press

the indicators in the evaluation system have not been empirically optimised through extensive expert consultation, and there is a lack of research on risk differences and adaptation paths in various disciplinary contexts. With the rapid development of technology, the forms and risks of AI literature retrieval tools will also be changing. In the future, research can be carried out in three directions: first, conduct large-scale tool tests and develop dynamic monitoring tools to improve the credibility assessment system; second, explore differentiated governance strategies according to the characteristics of different disciplines, such as humanities and social sciences, natural sciences, and medicine, and design corresponding credibility standards and usage guidelines; third, address the new challenges posed by multimodal AI retrieval and proactively advance research on the credibility governance of multimodal academic information, including images, data, and code.

Acknowledgments

The author acknowledges financial support from the Graduate Online Course Construction Project of Hangzhou Dianzi University (Grant No. ZXKC20240007).

References

[1] Dai, L. X., Tian, G. Q., and Liu, Y. N. International progress and lessons from the application of artificial intelligence in native electronic document management. Archival Studies, 2026, (4):132–141.

[2] Wang, X., Bai, Y. K., Ye, Y., et al. Artificial intelligence empowering scientometrics: Current status and future prospects. Library Forum, 2026, 46(8):1–14.

[3] Su, G., Yu, S. B., and Li, F. A credit risk early warning model based on public opinion. Insurance Research, 2021, (10):90–105.

[4] Jiang, S. H., and Zhang, K. Application and regulation of artificial intelligence in local legislation. Journal of Hunan University (Social Sciences Edition), 2019, 33(4): 139–146.

[5] Duan, H. L. Evolutionary risks and legal regulation of artificial intelligence search services. Journal of Shanghai Normal University (Philosophy and Social Sciences Edition), 2022, 51(2):100–108.

[6] Li, F. F. Current status and future prospects of large artificial intelligence models in open access resource platforms for philosophy and social sciences. Journal of Yangzhou University (Humanities and Social Sciences Edition), 2026, 30(4): 62–70.

[7] Yuan, P. B., and Bu, Y. F. AI search: Applications and risks of artificial intelligence in information retrieval. News Enthusiast, 2026, (5): 65–68.

[8] Wang, D., Wang, Y. M., Wang, H. J., et al. “Generative”or “traditional”? An empirical study on graduate students’ information retrieval behavior in the era of artificial intelligence. Library Science Research, 2025, (7): 79–91.

[9] Li, Q. F. AI-driven agent systems and knowledge production: Transformation, reshaping, and challenges. Chongqing Social Sciences, 2026, (7): 111–125.

[10] Zheng, H. A study on copyright infringement risks and compliance governance of library-developed knowledge bases based on retrieval-augmented generation (RAG) technology. Library Work and Research, 2026: (6), 29–42.

[11] Liu, C. Generative artificial intelligence and the development of digital archival resources. Library, 2025, (9): 39–46.

Copyright @ STEMM Institute Press