AI in Literature Reviews: a survey of current and emerging methods

1st Mohamed Saied
School of information technology and computer science
Nile university
Giza, Egypt
M.Abdelnaser2159@nu.edu.eg

2nd Nada Mokhtar
School of information technology and computer science
Nile university
Giza, Egypt
N.Hussain2128@nu.edu.eg

3rd Abdelrahman Badr
School of information technology and computer science
Nile university
Giza, Egypt
A.Ahmed2191@nu.edu.eg

4th Michael Adel
School of information technology and computer science
Nile university
Giza, Egypt
M.Hany2160@nu.edu.eg

5th Philopater Boles
School of information technology and computer science
Nile university
Giza, Egypt
P.Ayman2125@nu.edu.eg

5th Ghada Khoriba
School of information technology and computer science
Nile university
Giza, Egypt
ghadakhoriba@nu.edu.eg

2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) | 979-8-3503-6777-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/MIUCC62295.2024.10783597

***Abstract*—AI-assisted literature review tools have significantly enhanced researchers' productivity by streamlining the review process and reducing time consumption. Although traditional tools remain widely used, a new generation of tools leveraging state-of-the-art methods, including large language models (LLMs), is gaining popularity. However, LLMs face challenges such as hallucinations, which affect their reliability and accuracy. To address this, solutions such as knowledge augmentation are being explored. Additionally, combining knowledge-augmented LLMs with agentic frameworks has shown promise in improving their performance, making them more reliable and effective for literature review tasks.**

***Index Terms*—AI-assisted literature review, large language models, LLM hallucinations, knowledge augmentation, agentic frameworks.**

I. Introduction

The rapid advancement of artificial intelligence (AI) has fundamentally transformed many sectors, including healthcare, finance, education, and research. One of the growing areas of interest is using AI in academic research processes, particularly in automating literature reviews. The literature review is a critical component of academic research, as it provides an in-depth examination of existing studies, highlights gaps in knowledge, and identifies potential avenues for future investigation. However, as the volume of scholarly publications expands exponentially, researchers face increasing challenges in keeping up with the sheer quantity of new information. Manually curating, synthesizing, and critically analyzing various academic sources has become increasingly time-consuming and labor-intensive. Consequently, researchers and institutions have begun exploring AI-based tools and techniques to automate and streamline the literature review process.

Automation of literature reviews often referred to as "AI-assisted literature reviews," involves using machine learning algorithms, natural language processing (NLP), and other AI technologies to assist or entirely replace human involvement in certain stages of the review process. These technologies offer several potential benefits, including faster processing times, increased scalability, enhanced reproducibility, and reduced human bias. By leveraging AI tools, researchers can gather, sort, and synthesize relevant academic works more efficiently, enabling them to focus on higher-order cognitive tasks such as critical analysis and interpretation.

AI in automating literature reviews is still an emerging area of research, but it has shown promise in various applications. These include systematic literature reviews (SLRs), meta-analyses, scoping reviews, and narrative reviews across multiple disciplines. AI systems are being developed and refined to perform automatic citation extraction, document classification, topic modeling, summarization, sentiment analysis, and

network mapping functions. Machine learning models can also be trained to identify key concepts, themes, and relationships within large datasets of academic papers, helping researchers uncover new trends and synthesize data with greater accuracy and efficiency.

Despite its potential, integrating AI into the literature review process raises several critical questions and challenges. For instance, the issue of trust and reliability remains a significant concern. AI models rely on vast amounts of training data, and any biases or inaccuracies in these datasets can affect the quality and objectivity of their reviews. Furthermore, the interpretability of AI-generated insights is crucial; researchers must understand how and why the algorithms draw specific conclusions or summaries. In addition, there are concerns about the accessibility of AI tools, as their successful application often requires technical expertise and resources that may not be available to all researchers. Finally, ethical considerations must be carefully addressed, such as ensuring fairness and transparency in automated processes.

This literature review explores and evaluates the current state of AI in automating literature reviews. It examines the various AI technologies being employed, their advantages, and their limitations. By critically analyzing the progress and challenges in this field, this review aims to provide insights into the future potential of AI in academic research and offer guidance for researchers and practitioners looking to incorporate AI-driven methods into their literature review processes.

The following sections will review both established AI tools and emerging experimental systems. Finally, the paper will discuss the future directions of AI in literature review automation, considering how ongoing advancements in AI and machine learning could reshape academic research practices in the coming years. Through this comprehensive review, we aim to provide a foundation for understanding the current capabilities of AI in this domain and offer perspectives on its evolving role in the academic landscape.

II. Literature Review

A. Current methods

Scholars currently use many AI-assisted literature review systems to streamline the review process. Still, these systems vary in performance and scope and typically employ traditional machine learning methods. We will examine some of them below.

LitSuggest [1] concatenates different text fields (such as title, abstract, and keywords) and transforms them into a bag-of-words format for classification. It combines various classifiers like Ridge and Elastic Net, and their outputs are fed into a logistic regression classifier to produce the final classification. However, LitSuggest suffers from various limitations, such as only working with PubMed articles and not supporting pre-print servers like bioRxiv, only processing abstracts and not full-text articles, and accuracy can be compromised when abstracts are missing.

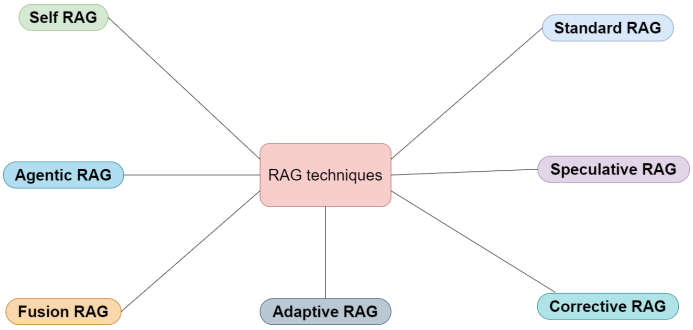

Fig. 1. Taxonomy of RAG techniques

Research Screener [2] uses paragraph embeddings for text representation. It is limited to a paper's title and abstract, limiting its scope to the early stages of the systematic review process.

These two tools can only be used in the screening phase; however, one tool that focuses on the extraction phase of the literature review is Dextr [3]. It is based on pre-trained embeddings (GloVe and ELMo), a bidirectional long short-term memory (LSTM) encoder, and a conditional random field. The model was developed and trained specifically on the methods sections of environmental health studies, allowing it to perform sequence tagging tasks by producing tags for each token in the input text from the methods section.

Overall, the most commonly used tools still lack the benefits of state-of-the-art techniques such as LLMs.

B. Emerging methods

Over the last few years, large language models (LLMs) have seen a meteoric rise due to their strengths in understanding context, generating human-like text, handling complex queries, and automating tasks such as summarization and translation. Their ability to adapt across domains and improve efficiency in various applications has driven widespread adoption. One such avenue of adoption is in assisting with literature reviews where they have shown state-of-the-art performance compared to traditional methods.

For example, Scite [4] takes advantage of the contextual information to prevent misinterpretations of citation intent, a common problem with traditional citation indices. The Scite tool addresses this gap by employing machine learning techniques to classify citations into three categories: supporting, contrasting, and mentioning. However, the classification of citation types may not capture all nuances, leading to potential inaccuracies in categorization, which is a current limitation of the tool.

Another example is Elicit [5], which summarizes essential information, including abstracts, interventions, outcomes, and participant details. It includes tools for brainstorming research questions and suggesting search terms. However, it relies on Semantic Scholar's database, which excludes licensed journals and paywalled content.

1) Knowledge augmentation: While LLMs provide many benefits compared to traditional methods in AI-assisted literature reviews, they have problems, particularly hallucinations. This affects the reliability and accuracy of the systems based on them. One way to address this issue currently being explored is augmenting LLMs with external knowledge. Fig. 1 shows a taxonomy of RAG techniques. Below is a review of some of the recently explored methods.

The KALA framework [6] enhances pre-trained language models (PLMs) by incorporating domain-specific knowledge during fine-tuning. This integration aims to improve task performance while maintaining efficiency and general knowledge. The framework has been tested across multiple datasets, including NewsQA and subsets of EMRQA for question answering, as well as various named entity recognition datasets. However, KALA's effectiveness is limited by the quality of external knowledge sources, which can impact overall performance.

The KAPING framework [7] introduces a knowledge-augmented prompting mechanism that utilizes knowledge graphs (KGs) for zero-shot question answering. The approach demonstrates superior performance compared to zero-shot baselines on KGQA benchmarks by retrieving relevant facts based on semantic similarity. Nonetheless, challenges remain in the retrieval scheme, multi-hop reasoning, entity linking, and evaluation metrics, which could benefit further refinement.

In [8], the authors address the limitations of traditional KGQA models, which are confined to the facts within knowledge graphs and cannot leverage common-world knowledge. Using a KGQA retriever based on the ReifKB and Rigel model families, the authors demonstrate that augmenting LM prompts with retrieved knowledge graph facts significantly enhances performance, achieving an average improvement of 83% over using an LM alone across four QA datasets. The method has limitations, including the absence of an integrated entity resolution system, reliance on annotated entities, and the need for retraining when new relations are added to the knowledge graph. Additionally, results are based on training and evaluating one dataset at a time, which may affect generalization.

The KAT model [9] integrates both implicit (commonsense) and explicit (external) knowledge to improve performance on vision-and-language tasks, particularly in the OK-VQA benchmark. Utilizing an encoder-decoder architecture, KAT effectively reasons over multimodal inputs. However, it faces challenges, such as the potential retrieval of generic or irrelevant information during explicit knowledge retrieval, which can introduce noise into the reasoning process. Additionally, relying on noisy or insufficient external knowledge sources may limit the model's reasoning capabilities.

In [10], the authors propose a framework that combines textual and visual relation knowledge to improve few-shot visual relation detection (VRD). The system utilizes pre-trained language models for textual knowledge and a visual relation knowledge graph for visual information. The framework enhances the generalization ability of VRD models but faces challenges in scaling to larger, more diverse datasets, mainly due to reliance on rule-based knowledge extraction.

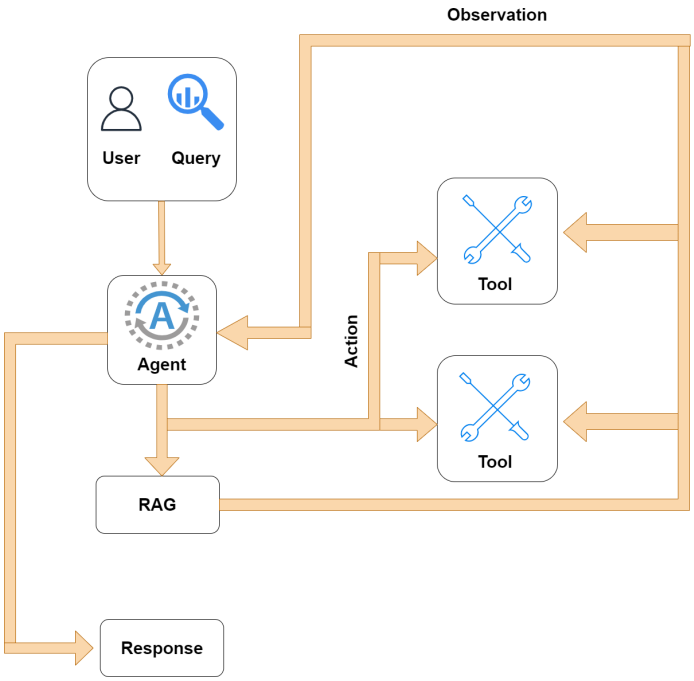

Fig. 2. Agentic RAG architecture

By integrating knowledge graphs, KAG [11] introduces a framework that enhances LLMs in professional fields, such as e-health and e-government. The system improves reasoning and task accuracy through domain-specific knowledge augmentation. Despite its effectiveness, the manual curation of knowledge graphs is labor-intensive and poses scalability issues, mainly when applied to diverse professional domains.

The Adaptive-RAG framework [12] presents a dynamic approach to question answering by selecting the most suitable retrieval strategy based on the complexity of the query. It operates across various datasets, including single-hop and multi-hop question-answering tasks. The limitations of Adaptive-RAG include the absence of dedicated datasets for training the query-complexity classifier, leading to potential mislabeling of queries due to reliance on model predictions and dataset biases. The current labeling approach, though effective, may still result in inaccuracies. Additionally, the classifier's architecture could be further improved.

2) Agentic frameworks: Agentic frameworks are models or systems that exhibit autonomy, self-directed decision-making, and task adaptability. Unlike traditional AI systems that follow preset rules or rely solely on static data, agentic frameworks empower AI to actively manage its behavior, dynamically choose actions, and interact with its environment in real-time. This approach enhances the AI's ability to solve complex problems by combining information retrieval, learning, and reasoning, allowing for more flexible, context-aware responses. Agentic frameworks are crucial in areas like autonomous systems, dynamic knowledge retrieval, and personalized user interactions, where adaptability and proactive decision-making

are key. Fig. 2 shows the architecture of Agentic RAG, which is one technique of RAG systems that uses agents. Below is a review of some notable works in this area.

In [13], the paper proposes an Agent Collaboration Network (ACN) framework designed to personalize AI search engines. The system uses multiple agents to retrieve and generate multimodal content and offer personalized responses. It incorporates user feedback through the Reflective Forward Optimization (RFO) method for real-time learning. However, it relies on synthetic datasets, which may not represent real-world scenarios accurately, and research gaps exist in evaluating AI search engines' responsiveness to user feedback.

KnowAgent [14] introduces a novel framework to improve planning in large language models (LLMs). It incorporates external action knowledge to reduce errors and hallucinations in reasoning tasks. KnowAgent has been tested on commonsense question answering and household tasks, but its effectiveness is limited to these domains. The manual creation of action knowledge bases is time-consuming and hinders scalability.

The authors in [15] present a hierarchical framework, Agentic Retrieval-Augmented Generation (RAG), for time series analysis. Multiple agents handle forecasting, anomaly detection, and classification tasks by retrieving prompts from historical patterns. While effective in managing distribution shifts and spatiotemporal complexities, the isolated operation of sub-agents restricts their performance in complex reasoning tasks. The framework's success is tied to the quality and efficiency of prompt retrieval.

DyLAN [16] introduces a dynamic framework for selecting and organizing LLM agents for tasks like reasoning and code generation. It uses an inference-time selection mechanism and an agent team optimization algorithm based on Agent Importance Scores. The framework is context-dependent and may struggle with unseen or novel tasks. Its reliance on LLM-powered rankers might also limit performance in unfamiliar scenarios.

In [17], the authors propose using LLM agents combined with Retrieval-Augmented Generation (RAG) to improve the accuracy of code search. The system, implemented in the RepoRift platform, enhances search precision by injecting relevant information into user queries. Although it improves Python code search, the system is limited to this language and heavily depends on internet retrieval, which may introduce outdated or irrelevant data.

In [18], the authors present a conceptual framework where multiple intelligent LLM agents collaborate, each with specific roles and capabilities, to enhance performance in tasks like artificial general intelligence (AGI). It includes dynamic agent addition, feedback, and supervisory mechanisms. However, scalability issues, potential security risks arise when managing many agents, and ethical concerns regarding autonomous decision-making in complex environments.

In [19], the authors introduce an AI-agent-based system designed to automate the process of systematic literature reviews (SLRs). It leverages LLMs to enhance the efficiency and accuracy of literature search, filtering, summarization, and analysis. However, the lack of comprehensive Boolean search strategies and unclear inclusion/exclusion criteria pose risks to search precision and review accuracy.

In [20] the authors introduce the LLM-Agent-UMF framework, which aims to seamlessly integrate active and passive core-agents with large language models (LLMs). The framework focuses on creating modular, scalable, and secure AI agents by differentiating the roles of LLMs and core agents. It emphasizes the modularity and reusability of agent-based systems, enhancing their flexibility for various applications. However, challenges arise with the complexity of synchronizing multiple active core-agent systems, which may affect scalability and maintainability. Additionally, ensuring effective communication between active and passive agents in complex configurations presents further obstacles.

III. Comparison between tools

Several LLM-based tools that can do literature reviews currently exist in the market. In Table I, we compare them across various features.

TABLE I
LLM-BASED TOOLS

Tool Name	**Results up till current day**	**Generates diagrams**	**Incorporates citations**	**Suggests ideas based on the review**
Consensus	×	×	✓	×
Scite	×	×	✓	×
Perplexity	✓	×	✓	×
Scispace	×	×	✓	×
Elicit	×	×	✓	×
Jenni.ai	×	×	✓	×
Textero.ai	×	×	✓	×
Samwell.ai	×	×	✓	×
OpenRead	✓	×	✓	×

IV. Open issues

Despite the existence of LLM-based tools that can help researchers with literature reviews, they still have limitations, such as giving results that are not up-to-date, their inability to generate diagrams and suggest ideas to be further explored based on the review of the papers as shown in Table I.

V. Conclusion

To conclude, AI-assisted literature review tools supercharge researchers' productivity and save them much time. While most of the tools researchers use still rely on traditional methods, many emerging tools are gaining momentum thanks to state-of-the-art techniques such as LLMs. However, LLMs suffer from hallucinations, which impact their reliability and accuracy. Solutions to this problem are being explored. One promising solution is augmenting LLMs with external knowledge. Moreover, combining knowledge augmentation with agentic frameworks provides an even better performance.

REFERENCES

[1] A. Allot, K. Lee, Q. Chen, L. Luo, and Z. Lu, "Litsuggest: a web-based system for literature recommendation and curation using machine learning," *Nucleic acids research*, vol. 49, no. W1, pp. W352–W358, 2021.

[2] K. E. Chai, R. L. Lines, D. F. Gucciardi, and L. Ng, "Research screener: a machine learning tool to semi-automate abstract screening for systematic reviews," *Systematic reviews*, vol. 10, pp. 1–13, 2021.

[3] V. R. Walker, C. P. Schmitt, M. S. Wolfe, A. J. Nowak, K. Kulesza, A. R. Williams, R. Shin, J. Cohen, D. Burch, M. D. Stout *et al.*, "Evaluation of a semi-automated data extraction tool for public health literature-based reviews: Dextr," *Environment international*, vol. 159, p. 107025, 2022.

[4] J. M. Nicholson, M. Mordaunt, P. Lopez, A. Uppala, D. Rosati, N. P. Rodrigues, P. Grabitz, and S. C. Rife, "scite: A smart citation index that displays the context of citations and classifies their intent using deep learning," *Quantitative Science Studies*, vol. 2, no. 3, pp. 882–898, 2021.

[5] J. Kung, "Elicit (product review)," *Journal of the Canadian Health Libraries Association / Journal de l'Association des bibliothèques de la santé du Canada*, vol. 44, no. 1, Apr. 2023. [Online]. Available: https://journals.library.ualberta.ca/jchla/index.php/jchla/article/view/29657

[6] M. Kang, J. Baek, and S. J. Hwang, "Kala: knowledge-augmented language model adaptation," *arXiv preprint arXiv:2204.10555*, 2022.

[7] J. Baek, A. F. Aji, and A. Saffari, "Knowledge-augmented language model prompting for zero-shot knowledge graph question answering," *arXiv preprint arXiv:2306.04136*, 2023.

[8] P. Sen, S. Mavadia, and A. Saffari, "Knowledge graph-augmented language models for complex question answering," in *Proceedings of the 1st Workshop on Natural Language Reasoning and Structured Explanations (NLRSE)*, 2023, pp. 1–8.

[9] L. Gui, B. Wang, Q. Huang, A. Hauptmann, Y. Bisk, and J. Gao, "Kat: A knowledge augmented transformer for vision-and-language," *arXiv preprint arXiv:2112.08614*, 2021.

[10] T. Yu, Y. Li, J. Chen, Y. Li, H.-T. Zheng, X. Chen, Q. Liu, W. Liu, D. Huang, B. Wu *et al.*, "Knowledge-augmented few-shot visual relation detection," *arXiv preprint arXiv:2303.05342*, 2023.

[11] L. Liang, M. Sun, Z. Gui, Z. Zhu, Z. Jiang, L. Zhong, Y. Qu, P. Zhao, Z. Bo, J. Yang *et al.*, "Kag: Boosting llms in professional domains via knowledge augmented generation," *arXiv preprint arXiv:2409.13731*, 2024.

[12] S. Jeong, J. Baek, S. Cho, S. J. Hwang, and J. C. Park, "Adaptive-rag: Learning to adapt retrieval-augmented large language models through question complexity," *arXiv preprint arXiv:2403.14403*, 2024.

[13] Y. Shi, M. Xu, H. Zhang, X. Zi, and Q. Wu, "A learnable agent collaboration network framework for personalized multimodal ai search engine," *arXiv preprint arXiv:2409.00636*, 2024.

[14] Y. Zhu, S. Qiao, Y. Ou, S. Deng, N. Zhang, S. Lyu, Y. Shen, L. Liang, J. Gu, and H. Chen, "Knowagent: Knowledge-augmented planning for llm-based agents," *arXiv preprint arXiv:2403.03101*, 2024.

[15] C. Ravuru, S. S. Sakhinana, and V. Runkana, "Agentic retrieval-augmented generation for time series analysis," *arXiv preprint arXiv:2408.14484*, 2024.

[16] Z. Liu, Y. Zhang, P. Li, Y. Liu, and D. Yang, "Dynamic llm-agent network: An llm-agent collaboration framework with agent team optimization," *arXiv preprint arXiv:2310.02170*, 2023.

[17] S. Jain, A. Dora, K. S. Sam, and P. Singh, "Llm agents improve semantic code search," *arXiv preprint arXiv:2408.11058*, 2024.

[18] Y. Talebirad and A. Nadiri, "Multi-agent collaboration: Harnessing the power of intelligent llm agents," *arXiv preprint arXiv:2306.03314*, 2023.

[19] A. M. Sami, Z. Rasheed, K.-K. Kemell, M. Waseem, T. Kilamo, M. Saari, A. N. Duc, K. Systä, and P. Abrahamsson, "System for systematic literature review using multiple ai agents: Concept and an empirical evaluation," *arXiv preprint arXiv:2403.08399*, 2024.

[20] A. B. Hassouna, H. Chaari, and I. Belhaj, "Llm-agent-umf: Llm-based agent unified modeling framework for seamless integration of multi active/passive core-agents," *arXiv preprint arXiv:2409.11393*, 2024.