

A Comprehensive Review on AI-Driven Literature Review, Retrieval-Augmented Generation, Knowledge Graphs, And Research Gap Detection

Under the Guidance of **Prof Monika S. Shirbhate**, **Prof Aparna R. Khaikar**

Riya Tatode (57), Anjali Raut (45), Manaswi Gawande (23), Trupti Bhongle (05)

Department of Information Technology, Prof. Ram Meghe College of Engineering and Management, Badnera, Amravati, Maharashtra, India | Academic Session 2026–27

I ABSTRACT

Literature review is a core task in academic and industry research, but the usual process takes major effort to find relevant papers, screen them, pull out information, compare results, trace citations, combine themes and spot research directions that remain unexplored. Recent progress in artificial intelligence, natural language processing, large language models, retrieval-augmented generation, knowledge graphs and autonomous agent workflows creates a chance to support this process. This paper reviews research related to IRIS (Intelligent Research & Information System), a proposed AI-based research assistant which aims to combine paper retrieval, scientific information extraction, summarization, citation-aware synthesis, knowledge-graph building, research-gap detection and structured report generation. We reviewed eighteen representative papers for five technical areas: AI- assisted systematic literature review (SLR), Scientific natural language processing (NLP), semantic search, RAG, agent reasoning and KN. We observed that existing approaches typically handle individual subproblems well but struggle with integration, provenance detection, long-context processing, explainability, systematic gap detection, etc. With these observations, we propose a conceptual IRIS framework. It consists of several domain-specialized autonomous agents agenting over a citation aware scholarly corpus, and a researchy knowledge graph. we summarize some evidence as to why IRIS is better considered as a human-in-the- loop system rather than a fully automated system of retrieving literature reviews, consistent with common understandings in research.

I. INTRODUCTION

A literature review combines the work that has been done on a topic to define the state of the knowledge in a research area. The traditional approach for a researcher browses the several scholarships sources looking for the relevant papers, reads the abstract and body texts of the papers, records the methods, results, compares the results against each other, and combines the results coherently. As the number of publication increases, the workflow becomes the challenging task to perform. In the seminar material for IRIS, the information overload, challenges of comparing papers and trends and a lack of support for extracting the structured knowledge are identified as the motivations for an intelligent research assistant [1]. The AI has increasingly used for the systematic literature review (SLR) processes. Bolaños et al. surveyed the AI-powered SLR approaches and illustrate that the existing systems can help the activities such as screening and extraction while the general research direction is the semi-automated literature reviews [1]. Han et al. investigated the retrieval-augmented generation for SLRs and discussed the automated retrieval, extraction, summarization, and trend identification [2]. The human-centered approaches have also demonstrated that the language models can help the researchers in developing the literature-review narratives rather than generating the summaries [3]. A useful review synthesizes the information while preserving the relations between the works, distinguishing the evidence and interpretation, highlighting the disagreements, and making the trail from the claims backed to the sources clear. The scientific NLP models such as SciBERT [4] and document-level representations such as SPECTER [5] provide the foundation for understanding and retrieving the scholarly

documents. The RAG [6], dense passage retrieval [7], and self-reflective RAG [8] provide the mechanisms for the grounding of the generated outputs in the retrieved evidence. The Knowledge graphs [9], [10] and GraphRAG [11] provide the structured approaches to representing and answering the corpus-level questions. IRIS is proposed to combine all these ideas in a workflow. The project seminar describes an architecture with the research-paper retrieval through the scholarly APIs, NLP and information extraction, AI summarization and paper analysis, research-gap and thematic analysis, knowledge-graph generation, and citation-aware report generation [1]. The goal is therefore not simply to generate the text, but to organize the scholarly evidence and support the researchers in understanding the research landscape.

A. Objectives of the Review

- Look into the studies that support retrieval and the ability to read papers.
- Look at the ways that RAG and agentic workflows can make the analysis of material using evidence better.
- Think about knowledge graphs as a way to show the connections between papers, ideas, methods and results.
- Identify the methodological gaps that result in the creation of the IRIS framework.

II. REVIEW METHODOLOGY

This paper contains a technology-oriented narrative literature review — and here's what makes it interesting — arranged according to the list of proposed IRIS components. . Initially, an extended list of related seminal and recent publications retrieved from academic databases. The review focuses on publications that have been peer-reviewed or represent research results in AI for Can't argue with that. No question. SLR scientific natural language understanding, semantic retrieval, RAG, self-reflection, agentic RAG, knowledge graphs, and graph-based retrieval. The final list of included works consists of eighteen items [1]-[18]. The reviewed papers were divided into five groups: (1) AI-assisted literature review,

(2) scientific language understanding. In this group, semantic retrieval falls, (3) — something often overlooked — retrieval-augmented generation, (4) reasoning and agentic workflows, and (5) knowledge-graph and graph-based retrieval. And, the division was based on the works' main contributions. Their relevance to the proposed IRIS solution, limitations reported by the

authors, or critical factors. Require consideration when designing an end-to-entire system for scientific literature reviews That's the bottom line. The review's purpose wasn't to provide an exhaustive list of all AI-driven literature-review technologies though to create. A sufficient body of evidence for the proposed project's technical feasibility and detail requirements for implementation and evaluation

III. LITERATURE REVIEW

A. AI-Assisted Literature Reviews

Bolaños et al. [1] offer one of the most relevant pieces of research for IRIS that I've seen.

Their survey of AI features — unexpectedly — literature review tools focuses on 21 le Human-AI Blend; significant AI. Impact different tools, most of support sureeg, wi a saller number supporting extraction. There is a clear distinction between the two, with the former greatly outnumbering the latter. The study demonstrates that there is a paradigm shift towards more AI-assisted and semi-automatic approaches compared to traditional tools. This, in turn, suggests that the functionalities of IRIS need to go beyond mere retrieval, screening, extracting, synthesizing, and maintaining traceability of all stages as interconnected processes. Not least, Han et al.From what I've seen, 2] concentrate on retrieval-augmented generation for automated systematic literature review tasks That changes things No getting around it. I think that Their review makes it clear that retrieval-augmented methods can be used to automate t processes of retrieval, summarization, extraction, and trend detection. To my mind, it is noteworthy that RAG, in particular, provides way of linking a language model to an external corpus of scholarly information, as opposed to relying only on information stored in weights of language model itself. From where I stand,The main finding of the paper is that quality of retrieval, computational cost are two critical factors that limit the practical application of retrieval-augmented approaches. Choe et al. That's huge. Still, [3] demonstrate how language models can help novice researchers That's the bottom line Can't argue with that. I find that Even so, their operate points to fact that literature review requires developing arguments and interacting with the literature, not just summarizing individual documents Simple as that. This human-centric approach has direct implications for IRIS in that intended platform should facilitate evidence-based inspection and comparison of papers to enable the researcher to reach independent conclusions

rather than presenting information as black box.

a. Scientific NLP and Semantic Retrieval

Scientific literature has a number of features that distinguish it from other domains, including a specific set of terms, relations, citations, mathematical expressions, and organizational structure. That's the reality. Beltagy et al. Honestly speaking, when you suggest about it, proposed SciBERT [4], which is a BERT-based language model pre-trained on scientific publications. From where I stand, Looking at this, SciBERT achieved positive results on several scientific natural language processing tasks, demonstrating the effectiveness of the domain adaptation process. To me, personally, I can use the same type of scientific representation in Iris for extracting concepts,, frankly speaking, methods, datasets, and findings from papers. No question. Cohan et al. To me, introduced SPECTER [5], a document-level representation model that uses citation information as a supervisory signal for representation learning. The results on document classification, citation prediction, and recommendation tasks show, and this is essential, that citation information can be used as a reliable signal for capturing scientific relationships between documents. For IRIS, this implies that similarity between documents is not limited to lexical overlaps found in the text of the documents; citation information can also be useful in identifying related works. That's huge. Semantic representations can be leveraged to jointly perform discovery and recommendation tasks. Karpukhin et al. When you think about it, presented Dense Passage Retrieval (DPR) [7], showing. I'd say that Dense representations allow for achieving strong performance in challenging open-domain question answering tasks.

c. Retrieval-Augmented Generation

Lewis et al. introduced RAG [6] as a framework I've always found that truth is, Their work demonstrated that retrieval can improve the quality of knowledge-intensive generation tasks and serves as a way of accessing external information. This is especially true for the IRIS project because generated summaries should be based on retrieved documents, not just on information contained within the model. Review claims in IRIS should be conditioned on the retrieved documents rather than relying solely on information contained within the model. To my mind, aI W Self-RAG by Asai et al Makes you think. I've noticed that What's more,

[8] — this is the crucial point — extends the retrieval framework by allowing the model to decide Pretty significant, honestly. Worth remembering. No kidding. On the flip side, from where I stand, When retrieval is useful and to reflect on both retrieved passages and generated content. The paper shows that factuality and citation accuracy can be improved in long-form generative tasks by conditioning generation on retrieved evidence. For IRIS, such a self-reflective turn would allow checking whether the generated statement has been supported. Personally, by a source, whether more evidence is needed, and whether a citation is appropriate. AI Liu et al. [12] highlight the limitations of long-context language models, namely the reduced ability to capture relevant information located in the middle of a long context Makes sense. To me, contexts tend to be used less often than information at the beginning or end Makes you think. This finding is especially relevant to literature reviews. When you suggest about it. That's the key. The way I see it, Since a large body of work is difficult to fit into the limited length of a prompt, I would argue that Iris should That's why, use retrieval, chunking, hierarchical summarization. Honestly speaking, structured representations rather than placing all papers directly into one context window. I've always found that honestly

d. Agentic and Reasoning-Based Workflows

Wei et al. [13] showed that prompting language models to generate chains of thoughts improved performance on complex reasoning tasks by encouraging the model to learn and explicitly represent reasoning steps during training. Although the technique is not specific to inquiry assistance, it nonetheless helps to think about complex scholarly tasks that require multi-step reasoning rather than single-step generation. To me, and, looking at this closely, Iris can set — and this is key — up this through explicit task decomposition, Like:. I've noticed that Retrieve, screen, extract, compare, identify themes, assess gap evidence, and draft Pretty significant, honestly. Singh et al Can't argue with that. [14] survey agentic RAG That's the bottom line. Honestly speaking, describe how autonomous agents can introduce planning,, which is interesting, reflection, tool use, and multi-agent collaboration into RAG pipelines. This is extremely relevant to the IRIS design because diverse agents can be employed to specialize — surprisingly — and this is key — in retrieval, paper

analysis, citation checking, knowledge graph construction, and gap detection Fairly straightforward, right? Agent specialization on top of that opens up opportunities for intermediate validation at various stages. The way I see it, The project seminar specifically proposes an agentic multi-AI workflow in. Multiple agents collaborate on retrieval, analysis, summarization, gap detection, and report generation [1]. The key analysis question isn't whether agents can generate text. Looking at this, whether agent coordination increases evidence coverage and decreases unsubstantiated claims. Worth remembering.

e. Knowledge Graphs and Graph-Based Retrieval

Hogan et al. [9] present an overview of knowledge graphs, including graph data models, knowledge extraction, quality assessment, and applications Makes you think. Ji et al. [10] survey representation learning, knowledge acquisition, completion, temporal graphs, and knowledge-aware applications. I find that Together, these works establish the basic principles for representing scholarly information in the form of entities and relationships. For IRIS, a research knowledge graph can be viewed as a collection of relevant entities such. Simple as that. As Paper, Author, Topic, Method, Dataset, Result, Citation, and Research Gap. Relations can include cites, authored-by, studies, uses-method, evaluates, supports, contradicts, extends, and addresses-gap. When you think about it, such a structure allows for storing and querying relationships between entities and facilitates graph-based exploration. Edge et al. [11] propose GraphRAG, which uses a graph index plus community summaries to retrieve relevant information from a large body of text. The paper demonstrates that GraphRAG achieves better performance than standard RAG approaches on global questions, including those concerning diversity and representativeness of the retrieved results. This is especially relevant to research questions such as "What are the main themes of this field?" or "What methods are most common, and what controversial findings have been reported?" A survey of automatic knowledge graph construction methods also shows that building a knowledge graph often involves knowledge acquisition, refinement, and evolution [15]. This, in turn, serves as a cautionary note for the IRIS project: a graph automatically constructed from papers may well contain noisy entities, ambiguous relationships, and outdated information. For this reason, graph curation

must be considered an integral part of the IRIS design.

iv. COMPARATIVE ANALYSIS

Table I shows how the surveyed studies fit the IRIS workflow. The comparison shows that most studies handle separate subproblems well, yet building a fully integrated workflow still poses a major engineering and research challenge.

Study	Area	Core contribution	IRIS relevance	Main limitation
Bolaños et al. [1]	AI for SLR	Examines how AI can help with screening and information extraction	Supports AI-assisted SLR activities	Does not provide a complete end-to-end workflow
Han et al. [2]	RAG for SLR	Explores retrieval, extraction, and summarization using RAG	Helps connect research papers with AI-generated analysis	Retrieval quality and computational cost can be challenging
Choe et al. [3]	LLM writing support	Studies how LLMs can assist researchers in preparing literature reviews	Encourages human-centered literature-review support	Does not cover the complete research workflow
SciBERT [4]	Scientific NLP	Uses scientific text to improve language understanding	Helps IRIS understand technical research content	Requires adaptation for specific research tasks
SPECTER	Scientific	Uses citation	Helps IRIS	Representation

[5]	retrieval	information to create meaningful paper representations	find papers with similar research themes	alone cannot perform the full review process
RAG [6]	Knowledge-intensive NLP	Combines retrieved information with language-model generation	Provides evidence-based generation for IRIS	Output quality depends strongly on retrieved information
DPR [7]	Semantic retrieval	Retrieves relevant passages using dense representations	Helps IRIS search by meaning rather than exact keywords	Mainly designed for general-domain question answering
Self-RAG [8]	RAG reflection	Allows model to retrieve, generate, and check its responses	Can improve the reliability and factuality of IRIS outputs	Adds additional complexity to the system
Hogan et al. [9]	Knowledge graphs	Explains how knowledge can be represented and connected through graphs	Provides a foundation for representing research relationships	Graph construction and maintaining data quality can be difficult
Ji et al. [10]	Knowledge graph survey	Reviews methods for building	Provides concepts and	Focuses on general knowledge

		and using knowledge graphs	methods useful for IRIS knowledge graph	edges rather than scholarly research
Graph RAG [11]	Graph-based RAG	Uses graph structures to understand information across a large collection	Can help IRIS identify themes and relationships between studies	Building and maintaining the graph may require significant effort
Liu et al. [12]	Long context	Examines how models handle information at different positions in long inputs	Supports the need for hierarchical processing in IRIS	Long-context processing can still reduce information utilization
Wei et al. [13]	LLM reasoning	Shows how step-by-step reasoning can support complex tasks	Provides a basis for breaking IRIS tasks into smaller stages	Not specifically designed for literature-review applications
Singh et al. [14]	Agentic RAG	Discusses planning, reflection, tool use, and agent collaboration	Supports the use of multiple specialized agents in IRIS	Agent coordination and evaluation remain challenging
KG construction survey	Automatic KG	Reviews approaches for acquiring	Provides guidance for	Automatic extraction may

[15]		and refining knowledge-graph information	building the IRIS research graph	introduce incorrect or duplicate relationships
PRISMA 2020 [16]	Systematic review reporting	Provides guidelines for making the review process transparent	Can help IRIS maintain a more organized and traceable review process	It is a reporting guideline rather than an AI implementation
LLM scientific assistance survey [17]	LLMs in research	Reviews how LLMs can support different scientific activities	Helps identify possible benefits and risks of using LLMs in IRIS	Covers a broad range of scientific applications
IRIS proposal [18]	Integrated research assistant	Combines retrieval, NLP, knowledge graphs, agents, and gap analysis	Provides the overall integrated workflow proposed for IRIS	complete framework still needs practical implementation and evaluation

V. RESEARCH GAPS IDENTIFIED

The literature shows several gaps which support IRIS

1) Fragmented Functionality = Due to of this, a researcher often has. Is interesting, to use several tools and move the information between them manually No getting around it. IRIS is proposed to reduce this difficulty by bringing the major stages of the review process into one workflow That's the reality. In my experience, I'd argue that

2) Limited Research-Gap Detection = Finding papers on a particular topic. No kidding. From where I stand, Is easier than deciding whether a real research gap exists No getting around it. Honestly speaking, sometimes the gap is related to a method that hasn't been tested with a particular dataset. Personally, iris Which means, looks at the information and relationships present surprisingly across reviewed papers to find possible research gaps.

3) Weak Evidence Traceability = Another issue is knowing whether an AI-generated statement is actually supported by a research paper. Then again, when you think about it, an answer generated by an AI model can sound correct even. When you think about it, when the supporting evidence is incomplete. Think about it. RAG and citation-aware approaches help to some extent Pretty noticeable, honestly. The source of an crucial statement should still be easy to identify. Personally, in IRIS, the generated information is intended to remain. Connected to the paper or passage from which it was obtained. Think about it. To me, this allows the researcher to go back to the source and verify the information.

4) Long-Context and Information Overload = A literature review may involve a major collection of papers. Simple as that Makes you think. Beyond that, reading all of their content

—. Looking at this closely, This is key — at the same time is difficult for both researchers and AI systems. And it shows. Liu et al That changes things. [12] discussed the problems language models —. Here's where it gets interesting — can face when relevant information is located within a long context. Because of this, IRIS shouldn't simply hand over all collected papers to an — and here's where it gets interesting — llm in one prompt And it shows. No question. Instead, the papers can be divided into — this is the. From where I stand, Crucial part — useful sections, relevant information can be retrieved first Simple as that. Summaries can be created at different levels. Honestly

speaking, the way I see it, The knowledge graph can then be used to connect information from different papers. Looking at this closely, Not only that ,

5) Lack of Standardized Evaluation = There's no single measure. No kidding. Can tell whether an AI literature-review system — and here's where it gets interesting — is working properly That's the bottom line. Honestly speaking, different parts of the system necessitate to be checked separately.

6) Human Oversight = Looking at this closely, This is Most of all, important That's the bottom line. When different studies have conflicting findings — and this is key — or when the evidence for a suggested research gap isn't satisfactory. Honestly speaking, when you think about it, AI-generated interpretations should also be checked. Before they're used in an academic report Worth keeping in mind. To me, meaning, IRIS is intended to work as a research-support — worth noting — system, helping the Pretty significant, honestly. From where I stand, Researcher handle large amounts of literature rather than replacing the researcher's judgment.

VI. PROPOSED IRIS FRAMEWORK

The architecture of the proposed IRIS framework is described in the project seminar [18]. The system has a research topic, question, or keywords as an input. The retrieval layer queries semantic databases like Semantic Scholar, arXiv, or Crossref through their API, saves the metadata and documents found, and prepares them for analysis by normalizing and deduplicating the data. The NLP layer extracts the relevant information such as title, authors, year of publication, research problem, methodology, dataset, results, limitations, and keywords. Scientific language models and embeddings may be used for semantic similarity and extraction. The extracted information is then structured as the evidence layer for further processing. The final dashboard allows searching for papers, applying filters, comparing, and exploring relationships in graphs, identifying themes, and generating a report with citations. The project seminar lists the described components as the key features of the IRIS framework [18].

A. Conceptual Workflow

User Query → Literature Search → Deduplication and Screening → Scientific NLP Extraction → Semantic Summarization → Knowledge Graph Construction → Cross-Paper Thematic Analysis → Research Gap Detection → Citation Verification → Review Report

and Interactive Dashboard

B. Research Gap Detection Strategy

A research gap detection algorithm should, first and foremost, find papers with similar topics and separate them into categories based on the methodology, dataset, and results. Second, it should identify under-represented combinations of topics, methodologies, datasets, and results, as well as disconnected areas in the graph. Third, a research gap should be found by comparing the limitations and future works sections of similar papers. Fourth, the algorithm should detect conflicting results and methodologies. Finally, the Gap Agent should suggest a research gap with supporting papers and a confidence score, and the researcher should validate the findings.

C. Citation-Aware Generation

The paragraphs produced by the system should be cited with the relevant evidence units. The citation manager can store the paper ID, URL, DOI, or the specific page/section, if available, with the claim. If there are not enough supporting items, then the system should not generate a citation and ask for the user's input instead.

VII. TECHNOLOGY STACK AND IMPLEMENTATION CONSIDERATIONS

The technology stack suggested by the project seminar includes React Makes you think. Js and Tailwind CSS are used for the frontend. Python with FastAPI is used for the backend. Transformer-based NLP is also part of the stack. When we look at this closely PyTorch is used for AI processing and PostgreSQL is used for data storage. Git and GitHub are used for version control [18]. These options are appropriate for a research assistant.

- A dashboard, a search function, filters, paper comparison capability, graph visualization and report preview are all part of the frontend features. That seems straightforward.

- Authentication, API orchestration, retrieval services, agent coordination, report generation and logging are all part of the backend features of the system. That changes things.

- Applications, in intelligence and natural language processing include scientific embeddings, information extraction, summarization, classification, evaluation of semantic similarity and verification of evidence.

- For storage relational metadata is kept in PostgreSQL. Honestly speaking graph representation is used via a

graph database or a graph layer. When you think about it

- For external academic sources, Semantic — surprisingly — scholar, arXiv, and Crossref are utilized; source licensing and API restrictions are applicable. The way I see it,

- Evaluation: a benchmark set. Has been carefully selected, consisting of research topics together with — something often overlooked — expert-labeled relevant papers, themes, citations, and candidate gaps.

VIII. EVALUATION FRAMEWORK

A proper evaluation of the project would entail comparing the IRIS results against the performance baseline of the procedures. The retrieval performance can be judged in terms of precision, recall, and nDCG@ ranks. In turn, the information extraction quality should be evaluated using precision, recall, and F1 scores measured against fields such as method, dataset, and finding. Meanwhile, the summary quality can be assessed according to their accuracy in terms of content coverage and supporting claims with citations. Finally, the citation generation performance should be evaluated against accuracy and relevance. Meanwhile, the research-gap detection component should be evaluated in terms of experts' judgment. Specifically, for research topics, the domain experts can assess whether the identified research-gap is novel, non-obvious, evidence-based, and actionable. Thus, it would be possible to report precision with evidence, annotated by the percentage of research gaps that would be agreed upon by the annotators, among other metrics. In turn, the time spent by the agent and the users' satisfaction should be measured, as the IRIS is designed to act as a research assistant.

Lastly, an ablation study should be conducted to measure the contribution of each component. For instance, the keyword-based search results can be compared against the search, RAG-based summaries, against standard document generation, flat database results against graph database results, and single-agent processing against multi-agent processing..

IX. CHALLENGES, ETHICS, AND LIMITATIONS

The use of AI to perform a literature review presents risks. I would argue that need to be addressed directly And this is what it boils down to. However, hallucinations could produce the fabrication of references. But, retrieval systems give a higher value to papers. Personally, are highly cited or easy to index and Which means, not highlight less prominent research.

Also, the various source APIs and publishers have various ranges of coverage, different licensing arrangements and different schedules for updating. Finally, the models have mimicked the biases present in the training data or in the scholarly record. IRIS should make provenance available, retain the source metadata, differentiate between evidence. It's as simple as that. From my position, Has been retrieved and that which has been interpreted and asks for human approval before any publication is complete. Meanwhile, a possible research gap cannot be presented as a proven scientific fact. Database terms and copyright should also be considered when storing or redistributing full-text papers. A limitation of the proposed framework is that it has not been tested experimentally as a complete system. Which is huge. The claims made in this paper are related to a design based on research rather than providing experimental evidence Worth remembering. In my experience, iris will surpass existing systems. When the framework is finally used, it should have information about. The datasets, model versions, retrieval settings, evaluation methods, error analysis, and repeatable experiments Worth considering. Can't argue.

X. FUTURE SCOPE

Further developments of IRIS can include a multilingual collection of scientific papers, multimodal comprehension of paper's content including graphs. Tables, trend analysis of the research over time, tailored recommendations depending on the field of a researcher, and a hypothesis generator. Looking at this closely, Indirect relationships between methods and outcomes can be found by extending the graph-based retrieval technique. I've noticed that Agent memory can enable long-lasting research to keep the same evidential basis through different working sessions. Big difference. Honestly speaking, an additional development may concern explainable research gap identification. Put differently, a user shouldn't receive one single statement about a gap. And, also should see the evidence path used by the algorithm, supporting papers, non-existing relationship, contradictory evidences. Probability of the recommended gap. And it shows. Because there are no explanations, the output will be more beneficial for academic supervision and won't be abused. Still, further development of the project can lead to PRISMA-oriented work flow. It implies that screening results, reasons of inclusion or exclusion, search dates. Evidence extraction can be recorded To be replicated by other researchers.

xi. CONCLUSION

This review has provided a review of research foundations related to. Plus, the IRIS, an Intelligent Research & Information System for assisted literature review. In my experience, important developments in AI-assisted screening. Extraction, scientific NLP, semantic retrieval, RAG, self-reflective generation, agentic workflows, knowledge graphs, and graph-based corpus analysis are highlighted in this literature. Makes sense. Nonetheless, integration, evidence traceability, long-context handling, validation of research gaps, and standard evaluation are significant problems. These limitations have been overcome through the proposed IRIS. Has connected scholarly retrieval, scientific information extraction, evidence-grounded summarization, knowledge graph creation, agentic multi-step analysis, research gap identification. Citation aware report generation. Personally, because of this, rather than any specific novel component, the main. From where I stand, Contributions of the proposed system are its integration process and human-in-the-loop approach. To me, for a final year engineering project, the next step will be to develop. Evaluate a tangible system based on certain clear benchmarks. The way I see it, Then again, as long as the proposed system can highlight accurate retrieval, faithful summary creation, traceable citations, useful thematic relations. No kidding. Verified research gaps suggested, the IRIS will be able to offer an effective platform for literature review.

REFERENCES

- [1] D. Scherbakov, N. Hubig, V. Jansari, A. Bakumenko, and L. A. Lenert, "The emergence of large language models as tools in literature reviews: A large language model-assisted systematic review," *Journal of the American Medical Informatics Association*, vol. 32, no. 6, pp. 1071–1086, 2025, doi: 10.1093/jamia/ocaf063.
- [2] S. Agarwal, G. Sahu, A. Puri, I. H. Laradji, K. D. Dvijotham, J. Stanley, L. Charlin, and C. Pal, "LitLLMs, LLMs for Literature Review: Are We There Yet?," *Transactions on Machine Learning Research*, 2025.
- [3] J. L. Lieberum, M. Toews, M.-I. Metzendorf, F. Heilmeyer, W. Siemens, C. Haverkamp, D. Böhringer, J. J. Meerpohl, and A. Eisele-Metzger, "Large language models for conducting systematic reviews: On the rise, but not yet ready for use—a scoping review," *Journal of Clinical Epidemiology*, vol. 181, art. 111746, 2025, doi: 10.1016/j.jclinepi.2025.111746.
- [4] A. Singh, A. Ehtesham, S. Kumar, T. T. Khoei, and A. V. Vasilakos, "Agentic Retrieval-Augmented Generation: A Survey on Agentic RAG," arXiv:2501.09136, 2025.
- [5] S. Guizani, T. Mazhar, T. Shahzad, W. Ahmad, A. Bibi, et al., "A systematic literature review to implement large language model in higher education: Issues and solutions," *Discover Education*, vol. 4, art. 35, 2025, doi: 10.1007/s44217-025-00424-7.
- [6] Y. Saleh, M. Abu Talib, Q. Nasir, and F. Dakalbab, "Evaluating large language models: A systematic review of efficiency, applications, and future directions," *Frontiers in Computer Science*, vol. 7, 2025, doi: 10.3389/fcomp.2025.1523699.
- [7] I. D. Mienye, N. Jere, G. Obaido, O. O. Ogunraku, and E. Esenogho, "Large language models: An overview of foundational architectures, recent trends, and a new taxonomy," *Discover Applied Sciences*, vol. 7, art. 1027, 2025, doi: 10.1007/s42452-025-07668-w.
- [8] F. Bolaños, A. Salatino, F. Osborne, and E. Motta, "Artificial intelligence for literature reviews: Opportunities and challenges," *Artificial Intelligence Review*, vol. 57, art. 259, 2024, doi: 10.1007/s10462-024-10902-3.
- [9] B. Han, T. Susnjak, and A. Mathrani, "Automating systematic literature reviews with retrieval-augmented generation: A comprehensive overview," *Applied Sciences*, vol. 14, no. 19, art. 9103, 2024, doi: 10.3390/app14199103.
- [10] K. Choe, S. Park, S. Jung, H. Kim, J. W. Yang, H. Hong, and J. Seo, "Supporting novice researchers to write literature review using language models," in *Extended Abstracts of CHI 2024*, 2024, pp. 1–9, doi: 10.1145/3613905.3650787.
- [11] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, "Lost in the middle: How language models use long contexts," *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 157–173, 2024, doi: 10.1162/tacl_a_00638.
- [12] P. Peng, Y. Zhu, Y. Liu, X. Bo, H. Shi, C. Hong, Y. Zhang, and S. Tang, "Graph Retrieval-Augmented Generation: A Survey," arXiv:2408.08921, 2024.
- [13] S. Gupta, R. Ranjan, and S. N. Singh, "A comprehensive survey of Retrieval-Augmented Generation (RAG): Evolution, current landscape and future directions," arXiv:2410.12837, 2024.
- [14] X. Hou, Y. Zhao, Y. Liu, Z. Yang, K. Wang, L. Li, X. Luo, D. Lo, J. Grundy, and

H. Wang, "Large language models for software engineering: A systematic literature review," *ACM Transactions on Software Engineering and Methodology*, vol. 33, no. 8, art. 220, 2024, doi: 10.1145/3695988.

[15] "Potential Roles of Large Language Models in the Production of Systematic Reviews and Meta-Analyses," *Journal of Medical Internet Research*, vol. 26, 2024, doi: 10.2196/56780.

[16] R. Raman, D. Pattnaik, L. Hughes, and P. Nedungadi, "Unveiling the dynamics of AI applications: A review of reviews using scientometrics and BERTopic modeling," *Journal of Innovation & Knowledge*, vol. 9, no. 3, art. 100517, 2024, doi: 10.1016/j.jik.2024.100517.

[17] B. Abu-Salih and S. Alotaibi, "A systematic literature review of knowledge graph construction and application in education," *Heliyon*, vol. 10, no. 3, art. e25383, 2024, doi: 10.1016/j.heliyon.2024.e25383.

[18] C. Peng, F. Xia, M. Naseriparsa, and F. Osborne, "Knowledge graphs: Opportunities and challenges," *Artificial Intelligence Review*, vol. 56, pp. 13071–13102, 2023, doi: 10.1007/s10462-023-10465-9.

[19] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, and H. Wang, "Retrieval-Augmented Generation for Large Language Models: A Survey," arXiv:2312.10997, 2023.

[20] Z. Zhang, M. Fang, L. Chen, M.-R. Namazi-Rad, and J. Wang, "How Do Large Language Models Capture the Ever-changing World Knowledge? A Review of Recent Advances," in *Proceedings of EMNLP 2023*, pp. 8289–8311, 2023, doi: 10.18653/v1/2023.emnlp-main.516.

[21] M. Salnikov, H. Le, P. Rajput, I. Nikishina, P. Braslavski, V. Malykh, and A. Panchenko, "Large Language Models Meet Knowledge Graphs to Answer Factoid Questions," in *Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation*, pp. 635–644, 2023.

[22] S. Verma, R. Bhatia, S. Harit, and S. Batish, "Scholarly knowledge graphs through structuring scholarly communication: A review," *Complex & Intelligent Systems*, vol. 9, pp. 1059–1095, 2023. Note: the article was published online in August 2022 but appears in the 2023 volume.

[23] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 494–514, 2022, doi: 10.1109/TNNLS.2021.3070843.

[24] M. Kejriwal, "Knowledge graphs: A practical review of the research landscape," *Information*, vol. 13, no. 4, art. 161, 2022, doi: 10.3390/info13040161.

[25] J. S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 494–514, 2022, doi: 10.1109/TNNLS.2021.3070843.