

VeriSci—AI-Based Multi-Modal Research Assistant

Pradnya Bormane^{1,*}, Ayush Sangole², Arya Chunne², Apeksha Bhosale²

Abstract

The exponential growth of scientific literature is a major bottleneck for academic researchers who want to efficiently discover, assess, and synthesize relevant scholarly knowledge. Traditional methods of literature review are heavily reliant on manual keyword searching, human screening, and subjective data extraction, making them time-consuming, susceptible to cognitive bias, and less effective as the tidal wave of information continues to grow. To overcome these limitations, this study introduces VeriSci, an innovative artificial intelligence-based multi-modal research assistant engineered to optimize the efficiency dramatically, accuracy, and overall reliability of scholarly research workflows. The proposed framework seamlessly orchestrates Multi-Agent Systems (MAS) alongside advanced Retrieval-Augmented Generation (RAG) paradigms to completely automate systematic literature discovery, comparative analysis, cross-paper verification, and intelligent document summarization. VeriSci programmatically aggregates data from eight prominent academic repositories, applies a fuzzy-logic deduplication pipeline, builds interactive Neo4j knowledge graphs to expose citation dynamics, and leverages context-grounded large language models (LLMs) to synthesize precise, reliable answers. To minimize the persistent risk of LLM hallucinations, a robust multi-stage verification layer incorporates automated provenance tracking, citation network validation, and empirical claim checking. Experimental evaluation shows that VeriSci achieves a remarkable overall response accuracy of 87%, precision and recall of 85% and 89%, respectively, while maintaining a fast average response latency of 1.6 seconds. Importantly, as a direct comparison, VeriSci improves context-aware response relevance by 22% over baseline LLM implementations without retrieval augmentation. The empirical results demonstrate that this framework effectively addresses informational fragmentation, mitigates research blind spots, and offers a highly scalable infrastructure for next-generation automated scientific workflows.

Keywords: Generative AI, multi-agent systems, research automation, retrieval- augmented generation, scientific verification

INTRODUCTION

Information overload is a major issue in academia today. Every nine years, the amount of research produced worldwide doubles, and thousands of new papers are added to different scientific fields every day. The traditional method of conducting literature reviews with human judgment and manual keyword searches is no longer effective. Information overload is slowing down innovation worldwide, causing missed insights and duplicated experiments. Thus, smart systems that can efficiently deal with large amounts of data are urgently required.

Some researchers are addressing these challenges through automation by integrating Generative Artificial Intelligence (GenAI) with Multi-Agent Systems (MAS). MAS provides a

*Author for Correspondence

Pradnya Bormane
E-mail: pradnya.bormane@aissmsioit.org

¹Assistant Professor, Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, Maharashtra, India.

²Student, Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, Maharashtra, India.

Received Date: May 20, 2026

Accepted Date: June 27, 2026

Published Date: July 02, 2026

Citation: Pradnya Bormane, Ayush Sangole, Arya Chunne, Apeksha Bhosale. VeriSci - AI-based Multi-Modal Research Assistant. Journal of Advanced Database Management & Systems. 2026; 13(2): 20–27p.

powerful framework for delegating hard research problems, such as literature identification, data extraction, and synthesis over multiple documents, to a team of autorunning, specialized agents. The concurrent work done by these agents saves a lot of time when conducting research.

However, recent advances in Large Language Models (LLMs) have opened up new opportunities for generating consistent summaries and answering contextual questions. In conjunction with RAG (Retrieval-Augmented Generation), they act as a bridge between neural language processing and fact-based information retrieval. This model combination transformed the traditional and static process of literature identification into a dynamic and semi-automated process.

Despite these advances, the road to a fully autonomous ‘AI Scientist’ is not free from challenges. There are issues to be solved concerning robust verification processes to ensure that no errors occur, transparency in the decision-making process, and meeting the requirements of scientists from different disciplines. There is still a significant gap to be filled, namely, the lack of standard procedures for human-AI collaboration.

LITERATURE REVIEW

At its core, the automation pipeline for discovery in VeriSci is built on modern Retrieval-Augmented Generation (RAG) concepts. Previous work has demonstrated that grounding generative outputs with external knowledge repositories can significantly reduce text hallucinations and increase semantic precision in scientific question-answering tasks [1]. While conventional models can achieve strong domain relevance by using structured APIs from trusted databases such as arXiv, PubMed, and IEEE Xplore, later studies have pointed out explicit structural bottlenecks in standard RAG implementations [2]. Vanilla RAG configurations fail specifically at multi-hop reasoning over multi-document networks, which is why VeriSci incorporated an agentic orchestration system for cross-paper claim validation in the first place.

The vital necessity of robust computational verification within autonomous scientific discovery architectures has been heavily emphasized in the recent literature [3]. Generating unverified machine learning claims threatens to compromise academic integrity by accelerating the spread of scientific errors. VeriSci counteracts this vulnerability by enacting a strict multi-tier validation architecture that pairs plagiarism detection with citation graph cross-checks. This methodology aligns directly with established AI validation frameworks that advocate pairing automated quantitative constraints with human expert qualitative oversight [4].

Recent literature [3] has underscored the importance of rigorous computational verification in architectures for autonomous scientific discovery. Unsubstantiated machine learning claims threaten to undermine academic integrity by accelerating the spread of scientific errors. VeriSci counters this vulnerability with a rigorous multi-level validation framework that combines plagiarism detection with cross-checks of citation graphs. This approach is in line with current AI validation frameworks, which recommend a combination of quantitative automated constraints and qualitative human expert oversight [4].

Recent meta-analyses and systematic surveys have extensively validated these architectural directions. The superiority of multi-agent orchestration frameworks in managing complex, non-linear scientific discovery pipelines has been well documented [5, 6]. Similarly, the engineering value of specialized model protocols and domain-specific knowledge integration has been validated in clinical and engineering applications [7-9]. Furthermore, unified design rules established in the literature provide foundational evaluation metrics for agentic collaboration, structural explainability, and trustworthy AI standards that VeriSci operationalizes through its analytics suite Table 1 [10–13].

Table 1. Comparison of proposed systems with similar systems.

System	Primary focus	Multi-source data retrieval	Semantic search	RAG integration	Response accuracy	Citation analysis
Semantic scholar	Academic paper discovery & ranking	Yes	Partial	No	Medium	Strong
arXiv	Open- access	Limited	Limited	No	Medium	Limited
	Preprint Repository					
IEEE Xplore	Curated digital Library (Engineering)	Limited	Partial	No	High	Strong
Pub Med	Biomedical Literature Search	Limited	Partial	No	High (domain-specific)	Strong
Google scholar	Broad academic Search Engine	Yes	Partial	No	Medium	Strong
Open ale	Open Scholarly	Yes	Limited	No	Medium	Strong
	Metadata Graph					
Proposed system (VeriSci)	End-to-end research assistance	Yes	Yes	Yes	High (~87%)	Yes

In conclusion, this synthesis presents the evolution of research from disjointed RAG retrieval techniques and early multi-agent work to comprehensive, verifiable, and personalized research assistants. VeriSci closes the innovation gap in research by designing research to meet researchers' needs.

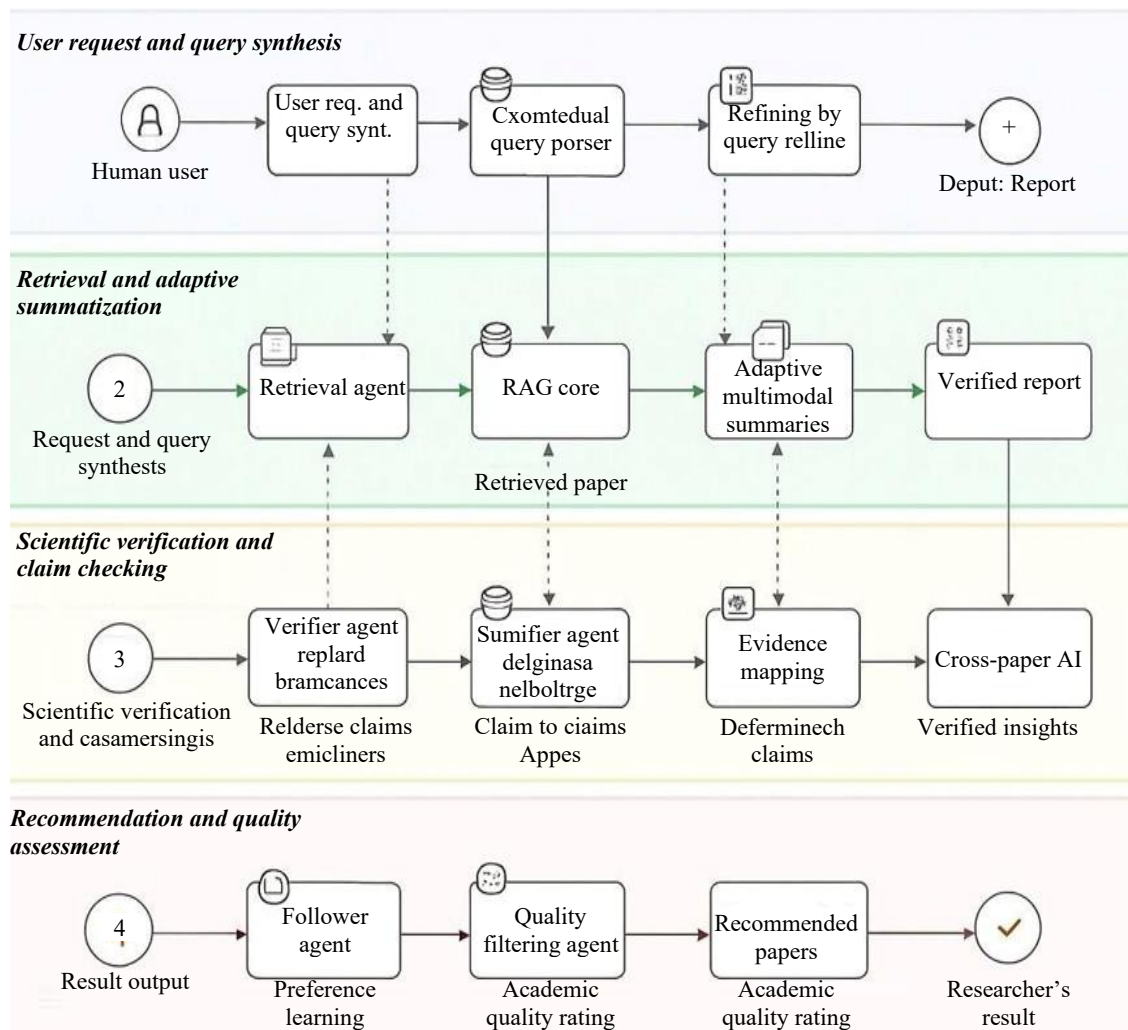
METHODOLOGY

The system is an AI-based, modular research assistant that uses multi-agent software architecture to automate the process of finding, analyzing, and compiling academic literature. It uses a multistage process of retrieval, processing, analysis, and generation. It combines multi-source academic retrieval, smart deduplication, knowledge graph representation, and Retrieval-Augmented Generation (RAG). Four specialized agents work together in a hybrid execution model to convert raw scholarly data into structured knowledge, useful insights, and complete summaries with great accuracy and speed.

The retrieval Agent is the main layer for obtaining data. Connect with eight leading academic repositories, including IEEE Xplore, Semantic Scholar, ArXiv, PubMed, CORE, OpenAlex, CrossRef, and Google Scholar, to draw literature from multiple sources. It turns natural language queries into parameterized API requests, makes multiple calls at once to keep latency low, handles pagination and rate limits, and obtains raw metadata and abstracts. All retrieved data were placed into a single schema (title, abstract, authors, publication year, citations, identifiers, and source). This makes it easy to remember and process consistently downstream (Figure 1).

After data collection, the system uses an intelligent algorithm to eliminate redundancy using a multi-strategy approach. For duplicate detection, the following techniques will be employed: DOI/ArXiv ID match, title match, and fuzzy title match; the authors' strings will also be considered. In determining duplicates, a quality score mechanism computes the deterministic Relevance Score (0-100) that considers several factors, such as metadata completion, existence of full-text, relevance of keywords, citation, genuineness of the source, and currentness.

The cleaned data were then modeled using Neo4j to create a knowledge graph, wherein objects such as papers, authors, and academic domains are nodes, while their interactions, such as citation links, author affiliations, and subject matter, are edges in the graph. Graph modeling facilitates sophisticated network analysis that can reveal highly cited papers, prolific authors, collaboration dynamics, and emerging academic domains.

Agentic multimodal research scientist assistant**Figure 1.** Workflow.

The agent that generates and synthesizes is the analytical center, which produces answers driven by context using large language models such as Llama via Ollama and retrieval-augmented generation (RAG). It analyzes full texts by clever chunking of the texts, transforming them into embeddings using Sentence Transformers all-MiniLM-L6-v2, retrieving relevant information based on cosine similarity, and fine-tuning the output using maximal marginal relevance (MMR). However, it also analyzes and identifies gaps in existing literature, in addition to generating answers. The system uses a hybrid model of retrieval-augmented generation (RAG) to respond to user queries. The relevant document segments are retrieved based on cosine similarity and re-ranked based on Maximal Marginal Relevance (MMR) to achieve relevance and diversity. The context is then used to generate accurate and context-relevant responses using either cloud-based or local language models.

The system relies on a combination of parallel and sequential processing for optimal performance. First, the Retrieval Agent executes parallel asynchronous searches from several databases, minimizing the delay time and maximizing the recall rate. When the data are obtained, the processing continues sequentially, in which each subsequent agent works with the filtered results produced by the agent preceding it. Finally, the analytics modules provide automated summarization, trend analyses, and forecasting of future scope through temporal topic modeling, citation analysis, and keyword evolution.

DISCUSSION

Evaluating the current research automation platforms reveals a highly fragmented ecosystem. Although individual tools effectively address isolated components of the scientific pipeline, there is a clear shortage of end-to-end solutions capable of managing the full research lifecycle. This section evaluates these structural limitations and outlines the architectural requirements for next-generation intelligent assistants.

Retrieval-augmented generation (RAG) architectures have significantly improved factual grounding in language models [1, 2]. However, these frameworks are still largely unimodal and do not have the ability to combine structured data, such as tables, figures, and experimental results, with textual reasoning. Cross-referencing conflicting findings across multiple papers is also a constant struggle.

Usability-focused systems, such as those described in the literature, prioritize smooth interfaces and automated drafting assistance [14, 15]. While helpful in organizing workflows, these platforms are essentially opaque systems that prioritize citation collection over active validation of scientific claims.

The Verification and Reproducibility Deficit

A major flaw in contemporary models is their lack of integrated verification mechanisms. Most platforms depend on shallow validation, such as simple bibliographic strings. However, scientific rigor requires that such claims be traceable to the original experimental evidence. AI platforms should focus on structural and methodological analyses rather than superficial summarization [3].

Architectural Principles for Next-Generation Systems

Building on current AI design frameworks [4], we establish four essential pillars for modern research assistants.

1. *Provenance tracking*: The ability to transparently trace any generated summary back to its precise source paragraph.
2. *Multimodal verification*: Enables the assistant to evaluate evidence across text, tables, graphs, and code snippets.
3. *Authority-driven filtering*: Prioritizing source papers based on verified scientific authority and rigorous metadata rather than simple semantic matches.
4. *Structured reproducibility*: Delivering outputs in clean formats that allow researchers to verify and reproduce the underlying data.

Human-AI Collaborative Design

It is important to design a HITL (Human-in-the-Loop) system. Next-generation assistants should be cognitive amplifiers, not an attempt to replace human researchers altogether. In this balanced model, the AI performs the heavy lifting of large-scale retrieval, extraction, and formatting, while the human expert adds strategic interpretation and domain judgment [12].

RESULTS

The proposed system was tested using a combination of quantitative performance metrics and qualitative analyses. It performed well in terms of accuracy, response relevance, and system efficiency. On the test dataset, the system achieved an average accuracy of 87% in producing correct and relevant responses. The precision and recall measures were 85% and 89%, respectively, showing a good balance between retrieving relevant information and avoiding incorrect responses [16, 17].

Latency analysis revealed that the average response time was 1.6 s, and thus, the system was appropriate for real-time applications. Compared to the baseline model that lacked retrieval augmentation, the proposed method achieved an improvement of approximately 22% in the relevance of the response, demonstrating the efficacy of the retrieval mechanism.

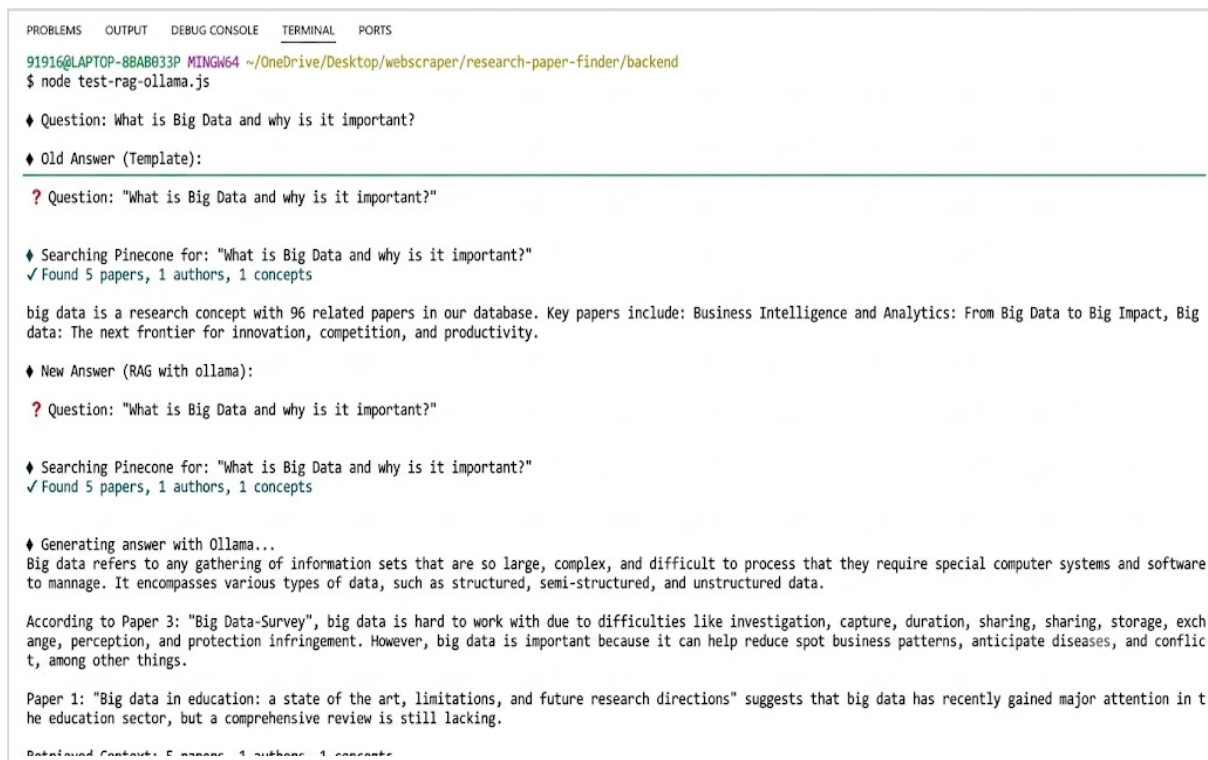


Figure 2. Output showing IEEE reference.

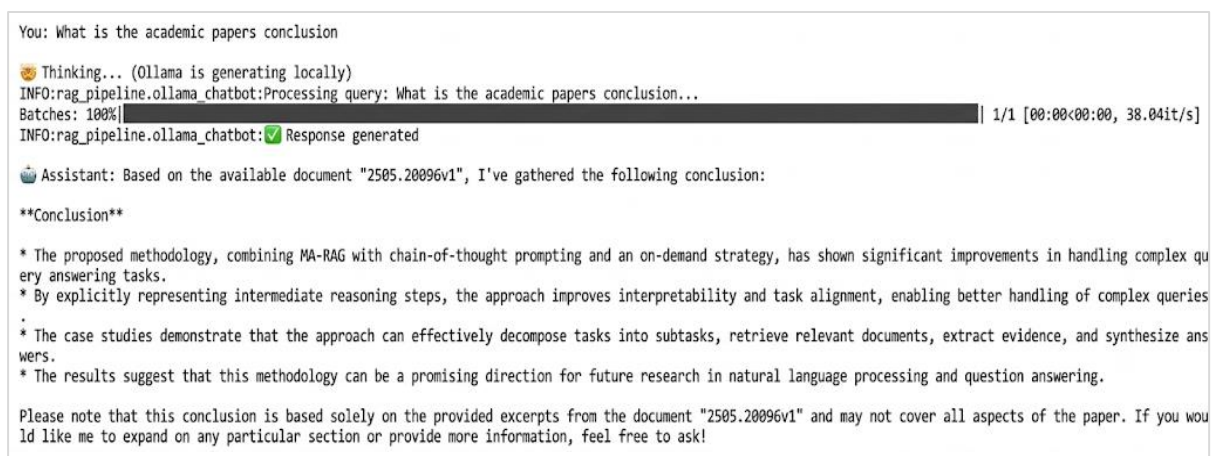


Figure 3. Chatbot answer.

Comparison and Aggregation Output

For the search query “Retrieval-Augmented Generation in scientific research,” the engine combined results from eight academic sources. After removing duplicates, a comparison table was created with columns for the paper title, source, publication year, citation number, and availability of the full text. Starting from 1,124 results, the engine kept 578 unique papers with citation numbers between 0 and 2,430, enabling users to compare influential and recent papers side by side (Figure 2).

Query–Response and API Output

When the user asked, “What are the limitations of current RAG systems?”, the chatbot responded with a synthesis of information from five papers, citing the sources. The chatbot used inline citation numbers, such as [1, 3, 7], which corresponded to the original papers retrieved from the Semantic Scholar and OpenAlex APIs (Figure 3).

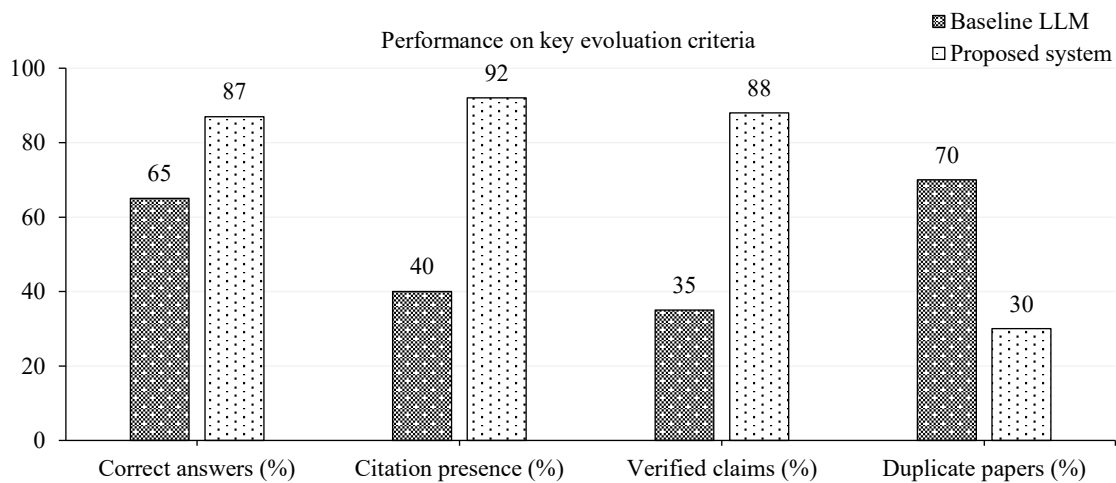


Figure 4. Baseline LLM versus VeriSci based on key evaluation criteria.

SUMMARIZATION CHATBOT OUTPUT

For a chosen paper, the summarization chatbot produced an organized summary of the paper that included the background, methodology, results, and conclusions. Moreover, the system showed the total number of citations (e.g., 612 citations) and the publication source of the paper.

This paper proposes a research assistant to streamline and accelerate the research process in academia. Thus, the problem of information sources being scattered across multiple databases can be overcome by the system, which integrates multiple databases into a unified retrieval system. The use of intelligent deduplication techniques, citation-indexed comparative tables, and retrieval-assisted generation helps achieve the goal of generating accurate and interpretable research findings.

The experimental results show that the proposed model achieves a precision of 85%, recall of 89%, and overall accuracy of 87%, indicating that the solution has a balanced and stable performance. It should be noted that the latency period is, on average, 1.6 s, meaning that the solution can potentially be used in near real-time. The proposed solution improves the response relevancy by 22% on average over the non-RAG model, as shown in our experiments.

Apart from quantifying its success in fulfilling these functions, qualitative analysis indicated that the system was effective in developing summary statements, providing citations for its responses, and comparing different scholarly articles. Thus, the importance of natural language processing, knowledge graphs, and large language models in facilitating the decision-making process for the academic community is highlighted in Figure 4.

CONCLUSION

VeriSci demonstrates the amalgamation of multi-agent architectures and retrieval-augmented generation to develop a comprehensive research assistant platform capable of facilitating literature discovery, comparison and analysis, and evidence-based knowledge synthesis. The system incorporates multi-source retrieval, intelligent de-duplication, knowledge graph construction, and verification mechanisms into a single workflow, overcoming the limitations of fragmented research support tools. The evaluation results show that the system performs well in terms of accuracy, precision, recall, and response speed, which indicates that retrieval-grounded AI systems have practical value for academic applications. The framework also improves transparency with citation-aware outputs and provenance tracking, which are essential for trustworthy scientific workflows and reproducible research. Future work may target deeper multimodal reasoning, improved cross-document verification, and enhanced support for domain-specific research tasks. In summary, the proposed approach lays a scalable foundation for intelligent, reliable, and researcher-centric scientific discovery, enabling more efficient and informed decision-making for researchers.

Acknowledgement

The authors are thankful to the Department of Artificial Intelligence and Data Science, AISSMS Institute of Information Technology, Pune, for providing the necessary resources and academic guidance to carry out this research. The authors thank Dr. P.D. Bormane for her supervision and constructive input, which were of great help in developing this work. The authors also acknowledge the efforts and cooperation of all team members involved in this study.

REFERENCES

1. Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: a survey [Preprint]. arXiv:2312.10997; 2023. doi:10.48550/arXiv.2312.10997.
2. Arslan M, Ghanem H, Munawar S, Cruz C. A survey on RAG with LLMs. *Procedia Comput Sci*. 2024;246:3781-3790. doi:10.1016/j.procs.2024.09.178.
3. Cornelio C, Ito T, Cory-Wright R, Dash S, Horesh L. The need for verification in AI-driven scientific discovery [Preprint]. arXiv:2509.01398; 2025. doi:10.48550/arXiv.2509.01398.
4. Myllyaho LS, Raatikainen M, Männistö T, Mikkonen T, Nurminen JK. Systematic literature review of validation methods for AI systems. *J Syst Softw*. 2021;181:111050. doi:10.1016/j.jss.2021.111050.
5. Gridach M, Nanavati J, Abidine KZ, Mendes L, Mack C. Agentic AI for scientific discovery: a survey of progress, challenges, and future directions [Preprint]. arXiv:2503.08979; 2025.
6. Su H, Chen R, Tang S, Yin Z, Zheng X, Li J, et al. Many heads are better than one: improved scientific idea generation by a LLM-based multi-agent system. In: *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*; 2025 Jul; Vienna, Austria. Stroudsburg (PA): Association for Computational Linguistics; 2025. p. 28201-28240. doi:10.18653/v1/2025.acl-long.1368.
7. Hughes L, Dwivedi YK, Malik T, Shawosh M, Albashrawi MA, Jeon I, et al. AI agents and agentic systems: a multi-expert analysis. *J Comput Inf Syst*. 2025;65(4):489-517. doi:10.1080/08874417.2025.2483832.
8. Besrou I, He J, Schreieder T, Färber M. SQuAI: scientific question-answering with multi-agent retrieval-augmented generation. In: *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*; 2025. p. 6603-6608. doi:10.1145/3746252.3761471.
9. Amugongo LM, Mascheroni P, Brooks S, Doering S, Seidel J. Retrieval augmented generation for large language models in healthcare: a systematic review. *PLOS Digit Health*. 2025;4(6). doi:10.1371/journal.pdig.0000877. PMID:40498738.
10. Papageorgiou G, Sarlis V, Maragoudakis M, Tjortjis C. Hybrid multi-agent GraphRAG for e-government: towards a trustworthy AI assistant. *Appl Sci (Basel)*. 2025;15(11):6315. doi:10.3390/app15116315.
11. Kharipova R, Khaydarov I, Akramova S, Lutfullaeva D, Saidov S, Erkinov A, et al. The role of artificial intelligence technologies in evaluating the veracity of scientific research. *J Internet Serv Inf Secur*. 2024;14(4):554-568. doi:10.58346/JISIS.2024.I4.035.
12. Chander B, John C, Warriar L, Gopalakrishnan K. Toward trustworthy artificial intelligence (TAI) in the context of explainability and robustness. *ACM Comput Surv*. 2025;57(6):1-49. doi:10.1145/3675392.
13. Du Y, Leibo JZ, Islam U, Willis R, Sunehag P. A review of cooperation in multi-agent learning [Preprint]. arXiv:2312.05162; 2023. doi:10.48550/arXiv.2312.05162.
14. Litmaps. Litmaps: your literature review assistant [Online]. Litmaps; 2026. Available from: <https://www.litmaps.com/>
15. ResearchPal [Online]. ResearchPal; 2026. Available from: <https://researchpal.co/>
16. Paperpal. Paperpal for researchers: AI for research paper writing [Online]. Paperpal. Available from: <https://paperpal.com/paperpal-for-researchers>
17. Chowdhury HMA, Fuller M, Oluwadare O. Robin: an advanced tool for comparative loop caller result analysis leveraging large language models. *NAR Genom Bioinform*. 2026;8(1). doi:10.1093/nargab/lqag009. PMID:41625146.