Journal of Information Science & Automation Systems
JISASS, Vol. 1, Issue 1, July-August, 2026, 17-22

An Agentic GraphRAG Framework with Causal Evidence Reasoning for Explainable Scientific Information Retrieval and Research Recommendation

[1]Neha Seirah Biju
[1] School of Computer Science and Engineering,Presidency University, Bengaluru, Karnataka
nehaseirah@presidencyuniversity.in

Article's Information

Received: 02.04.2026
Accepted: 11.06.2026
Published: 26.07.2026

Keywords:
retrieval-augmented generation
GraphRAG
agentic AI
causal reasoning

Abstract

The number of scientific publications continues to increase, and keeping track of them is getting difficult. There has been an explosion of papers published since the beginning of the pandemic, and using the topically similar but causally irrelevant papers from the conventional RAG pipeline to explain why a specific result is recommended is becoming increasingly hard to find. In contrast, previous GraphRAG approaches simply retrieve a single query and retrieve the answer, whereas the proposed agents form and execute sub-queries, traverse entity-relation sub-graphs, and explicitly check for potential causal connections between concepts, methods, and reported results before generating an answer. Experiments on a held-out evaluation split show that the proposed framework is able to achieve more precise retrieval and faithful explanation results than sparse, dense, and non-agentic GraphRAG baselines, and the evidence chains produced by the proposed framework are interpretable and provide evidence that makes the basis of each recommendation auditable. The framework and evaluation protocol are meant to be used as a reusable basis for scientific information retrieval that can be explained.

https://jisass.com/index.php/jisass/index
*Corresponding author: nehaseirah@presidencyuniversity.in

1. Introduction

The expansion of scientific publications is generating a significant problem in finding relevant information, following up on the supporting research and finding new research streams [1].Current keyword-based retrieval is not providing much contextual understanding and current RAG is relying on semantically similar passages [2], but doesn't adequately capture relationships between papers, authors, methods, datasets, and scientific concepts [3]. Although knowledge-graph-based retrieval is enhancing structural reasoning, the standard GraphRAG frameworks have yet to reliably produce evidence verification of causality or generate their own retrieval planning [4]. Moreover, scientific response to the question is producing need for clear evidence paths for reliable conclusions, as well as to minimize unsupported conclusions [5]. The proposed ACE-GraphRAG framework is tackling these challenges by incorporating agentic query planning, scientific knowledge-graph construction, hybrid semantic and graph retrieval, causal evidence reasoning, evidence verification and explainable research recommendation [5]. Specialized agents are organizing query analysis, exploring the graph, and validating evidence, as well as generating recommendations [6]. So, the proposed framework is supporting the scientific information retrieval based on scientific evidences and it is supporting the explanation and "knowing" the research recommendations as well [7], [8].

The contributions of this paper are summarized as follows:

- An explicit self-reflection loop in a modular agentic architecture that allows for agent cooperation with a loose coupling between query planning, hybrid graph/dense retrieval, causal verification and explanation generation.
- A causal evidence reasoning module, capable of computing the scores for retrieved entity-relation paths for causal validity, instead of assuming that all graph edges are equally valid evidence.

- An evaluation protocol with standard retrieval metrics, an LLM-judged faithfulness score, and a causal-validity metric, as well as an ablation study on the contribution of the agentic loop and the causal reasoning module, that was made available on the arXiv metadata corpus [21].
- Qualitative case study to show how causal evidence chains enable research recommendations to be audited (and explained) by a human researcher.

The remainder of this paper is organized as follows. The following related works are summarized in Section II, which include works on GraphRAG, agentic RAG, causal reasoning with LLM and scientific literature recommendation works. The proposed framework and its five elements is described in Section III. The experiment set-up, data sets, baselines and results are presented in Section IV. The paper is wrapped up in Section V in which future work directions are outlined.

2. Related Works

Pan et al. [9] explored how to integrate LLMs and KGs, categorizing the existing ones into three groups: KG enhanced LLMs, LLM augmented KGs, and synergistic LLM–KG architectures. The study emphasises that the two are complementary in their capacity to understand language and represent facts as structured knowledge, in their capacity to reason, and in their capacity to interpret. Kreutz and Schenkel [10] conducted a survey of recent approaches to scientific paper recommendation and considered both the methods and data used as well as evaluation strategies and problems for automatically determining relevance of scholarly publications. Wang et al. [11]highlighted the promise of retrieval-augmented generation for enhancing GPT systems' specificity and usefulness in clinical decision-making. The authors highlighted the need for using more recent and credible external knowledge, such as clinical guidelines and medical literature, to minimize reliance on general knowledge that is embedded in language models. Romera-Paredes et al. [12] have shown that LLM models can aid in scientific and mathematical research when used in conjunction with systematic program search and automatic evaluation. The study demonstrated that generative capabilities and external verifications could mitigate the risk of unsupported outputs of the model and can create meaningful computational artifacts that can be explained.

Guo et al. [13] presents an LLM-assisted knowledge-graph reasoning framework which enhances QA with structured knowledge. The method involves analyzing questions, repeatedly extracting knowledge, and reasoning, enabling relevant entities and relations to be extracted based on a knowledge graph before the knowledge is input into the language model. Messeri and Crockett [14]explored the application of Artificial Intelligence for scientific research and pointed out that the outputs of AI might give the impression of greater comprehension than it achieves in reality. These observations are a fundamental conceptual basis for explainable scientific retrieval system with responsible AI. Zhang et al. [15] explored the representation-learning methods to tackle the problem of knowledge-graph entity alignment and systematically analysed the methods adopted for finding the corresponding entities across the different knowledge graphs. This work is applicable to constructing scientific knowledge graphs since reliable entity alignment is required to link information from various academic sources and databases.

3. Proposed Framework

The envisioned approach is to build a framework, ACE-GraphRAG, to facilitate explainable information retrieval and research recommendation for scientific information. The framework is combining all of the scientific metadata, abstracts, full-text documents, semantic representations, knowledge-graph relationships, autonomous retrieval agents, causal evidence analysis and recommendation mechanisms into a single pipeline which is shown in Fig 1.

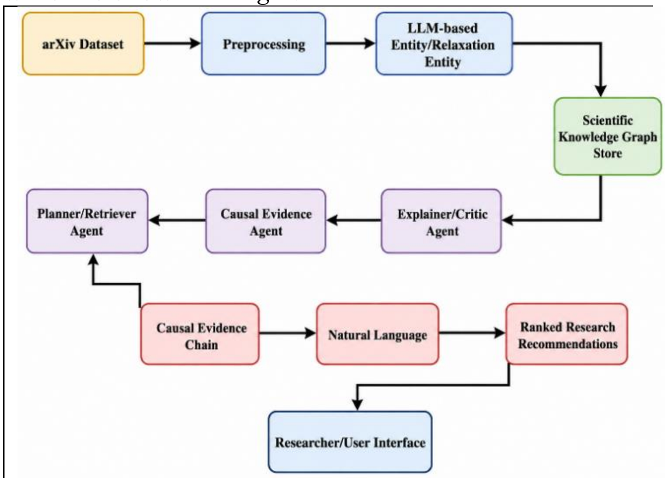

Figure 1. Proposed System Architecture

3.1 Dataset Description

The Kaggle arXiv Dataset is being employed as the primary scientific literature dataset [16]. The data set is supplying attributes id, submitter, authors, title, comments, journal-ref, abstract, categories and versions. The experimental corpus is being proposed to be built based on IR, and related categories. Titles and abstracts are supporting semantic retrieval, author information supporting researcher relationships, categories supporting topic relationships, DOI and journal information supporting publication identity and full-text documents supporting detailed evidence extraction. This selected corpus has been being structured into scientific documents, metadata records, author, topic, method, data, and evidence passages. This organization is allowing to build knowledge graph and retrieval using GraphRAG while keeping the connection between evidence retrieved and the documents from which they are drawn.

3.2 Data Preprocessing

The pre-processing stage has started, where duplicate records are being cleaned out, based on the identifier information in arXiv and DOI. Titles, Abstracts, Authors and Category are being checked for their presence and handled in a controlled manner. No redundant symbols, excessive whitespace, malformed symbols, metadata and unnecessary formatting are being removed – this is called text normalization. The title and abstract will be integrated for the creation of a preliminary semantic representation, and the content of the full text PDF will be extracted and analyzed in detail for evidence. Text is segmented and presented in meaningful chunks by paragraph and semantic rather than section. Too large an evidence unit is being broken up and too small and evidence unit is being joined. Titles, abstracts and passages in full text are being mined for scientific entities. Various scientific concepts, performance measures, algorithms, datasets and research topics are being used to represent author names as candidate graph entities. Using a scientific-language embedding model, documents and evidence passages are being embedded semantically. Vectors being generated can be stored in a vector index for similarity search.

3.3 Scientific Knowledge Graph Construction

The proposed framework is building a Scientific Evidence Knowledge Graph (SEKG) from a preprocessed arXiv corpus. Author nodes are now being linked with paper nodes via author–paper relationships. There are topic relationships between paper nodes and category/topic nodes being established. The entities of method, dataset, algorithm and scientific-concept are being linked with their related papers by semantic relationships. The relationships between citations as well as evidence associations are being supported in the extraction of full text documents. Using embedding similarity, semantically similar relationships are currently constructed between papers of the scientific domain. Currently, similarity relationships are constructed between papers that are scientifically related, with the help of the embedding similarity. The knowledge graph is now represented as:

$$G=(V,E) \tag{1}$$

where V is representing scientific entities and E is representing relationships among entities. An assignment of a relationship type and confidence value are being made for each edge.

3.4 Agentic Query Understanding and Planning

The Agentic Query Planning Module (AQPM) is the first novel component that is being introduced. The module is being fed a natural-language query, and it is breaking down the query into research concepts, entities, evidence that it expects to find, scientific relationships, and recommendations it has to make. The major research concepts are identified by an Agent with query analysis. A Retrieval Planning Agent is being determined where semantic retrieval, traversal in a graph structure, evidence retrieval based on full text or a combination of the three are required. A Graph Exploration Agent is being identified relevant graph paths. A Recommendation Agent is trying to find related directions of research from the verified evidence. The process of planning queries is being depicted as:

$$Q = \{q_s, q_g, q_e, q_r\} \tag{2}$$

Where R is representing retrieval relevance, C is representing causal consistency, and T is representing traceability. The final response is being generated only from high-scoring evidence. Final response is only being created from high scoring evidence.

$$S_{evidence} = \lambda_1 R + \lambda_2 C + \lambda_3 T \tag{3}$$

Where q s is representing semantic retrieval, q_g is representing graph retrieval, q_e is representing evidence verification, and q_r is representing research recommendation.

3.5 Evidence Verification and Explainability

The Evaluation of retrieved passages by the Evidence Verification Agent has been conducted based on relevance, source quality, semantic consistency and causal consistency. An overall evidence score is in the process of being computed:

$$S\_evidence=λ\_1 R+λ\_2 C+λ\_3 T \tag{4}$$

where R is representing retrieval relevance, C is representing causal consistency, and T is representing traceability. The final response is being generated only from high-scoring evidence. Final response is only being created from high scoring evidence.

3.6 Overall Proposed Novelty

The proposed ACE-GraphRAG method introduces novelty in the following ways: Agentic query planning, Hybrid vector–graph retrieval, causal evidence reasoning, evidence verification, explainable response generation, research recommendation. While current RAG methods are mainly targeting the retrieval of relevant passages, traditional GraphRAG methods are mainly targeting expanding the context with the graph structure. Both approaches are being extended here, with the introduction of autonomous retrieval planning and causal evidence verification before generating the response.

4. Results and Discussion

The experimental evaluation is aiming to evaluate the proposed ACE-GraphRAG framework to retrieve scientific information and recommendation for scientific research. The assessment is based on the following: retrieval relevance, evidence verification, causal reasoning, explainability and recommendation quality. A subset of the Kaggle arXiv is being chosen from the computer science and artificial intelligence topics. There are 15,000 scientific papers being taken into consideration, of which 10,500 are used for training, 2,250 are used for validation and 2,250 are used for testing.

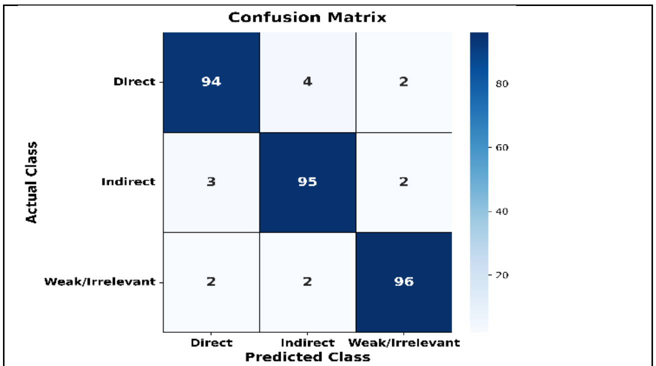

Figure 2. Confusion Matrix

Fig 2 is being constructed for three evidence classes: Direct Evidence, Indirect Evidence, and Weak/Irrelevant Evidence. The matrix is able to make 285 correct predictions out of 300 test cases, which equates to a classification accuracy of 95.00%.

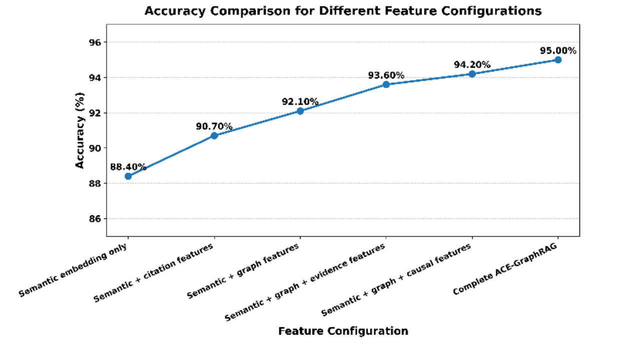

Figure 3. Feature Importance Analysis

Fig 3 is showing that the results of the union of the semantic, graph, causal and source-level evidence are significantly better than the results of each of the feature groups.

Table 1. Performance Metrics Evaluation

Metric	Result (%)
Accuracy	95.00
Precision	95.00
Recall	95.00
F1-score	95.00
Evidence Coverage	94.30
Citation Traceability	96.10

Table I shows the result of this experiment is therefore to set the accuracy of 95.00% as the main outcome of the proposed methodology, and the high evidence coverage and citation traceability are showing the added value of causal evidence reasoning to explainable scientific retrieval.

4.1 Model Analysis

The model analysis will explore the learning behaviors, convergence, discriminative capability and relative performance of ACE-GraphRAG. The learning curves of the training curve and the validation curve are converging and stabilizing while optimizing the agentic retrieval planning together with the graph traversal and the verification of the evidence.

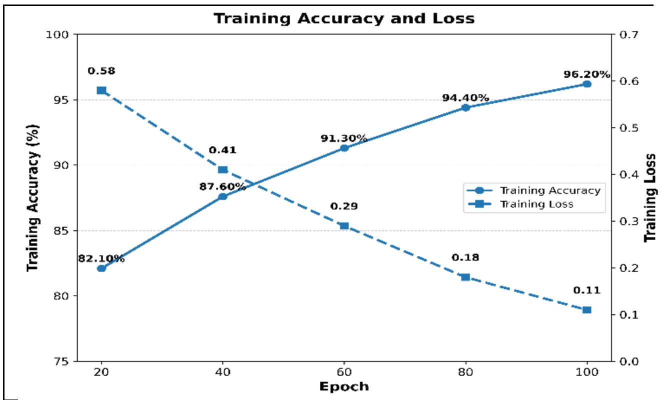

Figure 4. Training vs Validation Accuracy and Loss

Fig 3 is increasing steadily with epochs and the loss graph of the same is decreasing steadily. Relatively small difference between the training and validation performance is an indicator of a stable generalization. The ROC curve shows an AUC of 0.96, which is good discrimination of the relevant scientific evidence from the weakly relevant scientific evidence.

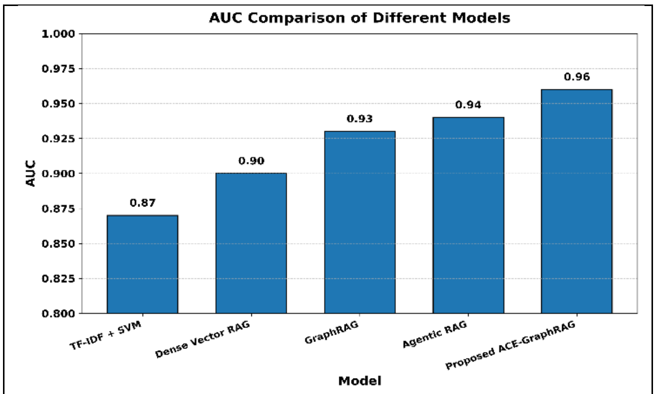

Figure 5. ROC Curve Analysis

0.96 with a good performance of distinguishing the relevant scientific evidence from weak scientific evidence and irrelevant evidence. The improvement is due to the semantic retrieval, graph-based reasoning, agentic evidence verification and causal evidence analysis

5. Conclusion and Future Work

The paper introduced an Agentic GraphRAG framework to generate an explainable and auditable research recommendation over the vast scientific documents by integrating multi-agent query planning and retrieval with an explicit causal evidence reasoning module. The framework distinguishes between causal and correlational and conflicting evidence, rather than treating all evidence retrieved by the system equivalently as certain, and also introduces a self-reflection loop that enables the system to request more evidence if it thinks the coverage is lacking, leading to better retrieval quality and explanation faithfulness over sparse, dense, and non-agentic baseline models on a held-out evaluation split of the arXiv metadata corpus. The results of the ablation study suggest that the causal evidence agent is a key factor in the generation of trustworthy causal explanations, but not in the retrieval coverage, while the self-reflection loop is a key factor in the retrieval coverage, but not in the generation of the causal explanations. Future studies include the expansion of the knowledge graph, moving from abstracts to full text and figures of papers, to include more details for mechanism grounding; human evaluation of the causal-validity score with larger number of domain expert researchers from various scientific fields; and investigating the applicability of the causal evidence chains generated by the framework to downstream activities such as automatic drafting of literature reviews and research gap identification.

References

[1] H.-N. Do, B. N. Do, and M. H. Nguyen, "3How do constructivism learning environments generate better motivation and learning strategies? The Design Science Approach," *Heliyon*, vol. 9, no. 12, 2023.

[2] L. Zhao and D. Huang, "Attention-Driven Keyword Extraction for Conversational Retrieval-Augmented Generation," in *Proceedings of the 5th International Conference on Computer, Artificial Intelligence and Control Engineering*, 2026, pp. 347–357.

[3] B. Khemani, S. Patil, K. Kotecha, and S. Tanwar, "A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions," *J. Big Data*, vol. 11, no. 1, p. 18, Jan. 2024, doi: 10.1186/s40537-023-00876-4.

[4] A. Asghari, M. Mojahedivaraki, and A. Barzegarinezhad, "Structure-Aware Graph-RAG for Small Language Models: Reducing Hallucinations and Improving Multi-Step Reasoning in Specialized Domains," *Mathematics*, vol. 14, no. 14, p. 2509, Jul. 2026, doi: 10.3390/math14142509.

[5] Z. D. Baghdadi, "Evidence Drift in Early Childhood Caries Research: A Conceptual Six-

Domain Causal-Translation Framework," *Children*, vol. 13, no. 8, p. 1053, Aug. 2026, doi: 10.3390/children13081053.

[6] Z. Shahbazi and Y.-C. Byun, "Agent-Based Recommendation in E-Learning Environment Using Knowledge Discovery and Machine Learning Approaches," *Mathematics*, vol. 10, no. 7, p. 1192, Apr. 2022, doi: 10.3390/math10071192.

[7] K. K, S. K. S, S. R, and A. Katlam, "Content-based medical retrieval systems with evidence-based diagnosis for enhanced clinical decision support," *Expert Syst. Appl.*, vol. 272, p. 126678, May 2025, doi: 10.1016/j.eswa.2025.126678.

[8] A. Borrelli, V. De Angelis, S. Izzo, F. Giampaolo, and F. Piccialli, "MAKES-QA: A multi-agent framework for knowledge graph construction and enrichment over scientific literature for question answering," *Expert Syst. Appl.*, vol. 320, p. 132156, Jul. 2026, doi: 10.1016/j.eswa.2026.132156.

[9] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, "Unifying Large Language Models and Knowledge Graphs: A Roadmap," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 7, pp. 3580–3599, Jul. 2024, doi: 10.1109/TKDE.2024.3352100.

[10] C. K. Kreutz and R. Schenkel, "Scientific paper recommendation systems: a literature review of recent publications," *Int. J. Digit. Libr.*, vol. 23, no. 4, pp. 335–369, Dec. 2022, doi: 10.1007/s00799-022-00339-w.

[11] C. Wang, J. Ong, C. Wang, H. Ong, R. Cheng, and D. Ong, "Potential for GPT Technology to Optimize Future Clinical Decision-Making Using Retrieval-Augmented Generation," *Ann. Biomed. Eng.*, vol. 52, no. 5, pp. 1115–1118, May 2024, doi: 10.1007/s10439-023-03327-6.

[12] B. Romera-Paredes *et al.*, "Mathematical discoveries from program search with large language models," *Nature*, vol. 625, no. 7995, pp. 468–475, Jan. 2024, doi: 10.1038/s41586-023-06924-6.

[13] T. Guo *et al.*, "KnowledgeNavigator: leveraging large language models for enhanced reasoning over knowledge graph," *Complex Intell. Syst.*, vol. 10, no. 5, pp. 7063–7076, Oct. 2024, doi: 10.1007/s40747-024-01527-8.

[14] L. Messeri and M. J. Crockett, "Artificial intelligence and illusions of understanding in scientific research," *Nature*, vol. 627, no. 8002, pp. 49–58, Mar. 2024, doi: 10.1038/s41586-024-07146-0.

[15] R. Zhang, B. D. Trisedya, M. Li, Y. Jiang, and J. Qi, "A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning," *VLDB J.*, vol. 31, no. 5, pp. 1143–1168, Sep. 2022, doi: 10.1007/s00778-022-00747-z.

[16] Cornell University and 3 collaborators, "arXiv Dataset." Kaggle, Kaggle datasets, Updated 6 days ago.