SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Retrieval-Augmented Scientific Reasoning Framework Using Knowledge Graphs And Large Language Models

Saranya Sampathkumar[1], Suresh Arumugam[2], Mohana Thiruchenduran[3], Jyotsna Suryavanshi[4], Ochilova Farida Bakhriddinovna[5], Dr. M. Abirami[6], Rakesh Arya[7]

[1] Assistant Professor, Department of Civil Engineering, Sona College of Technology, Salem-5, Tamil Nadu, India. Email: saranyasona07@gmail.com ORCID: 0000-0002-7497-0427
[2] Scientist, Central Research Laboratory, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Enathur, Kanchipuram, Tamil Nadu 631552, India. Email: Suresh@maher.ac.in
[3] Associate Professor, Department of Biochemistry, Meenakshi Ammal Dental College and Hospital, Meenakshi Academy of Higher Education and Research, India. Email: mohanat@maher.ac.in
[4] Department of Engineering, Science and Humanities, Vishwakarma Institute of Technology, Pune, Maharashtra 411037, India. Email: jyotsna.suryavanshi@vit.edu
[5] Assistant, Department of Pedagogy and Psychology, Samarkand State Medical University, Samarkand, Uzbekistan. Email: faridaochilova90@gmail.com ORCID: 0009-0000-0204-9041
[6] Assistant Professor, Department of Computer Science and Engineering, Panimalar Engineering College, Tamil Nadu, India. Email: abiprj@gmail.com ORCID: 0000-0003-4935-8938
[7] Assistant Professor, Department of Computer Application, Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand, India. Email: pc.mtech.mca@dbuu.ac.in ORCID: 0000-0003-2436-3060

Abstract
The proliferation of scientific literature exponentially has rendered it more challenging to researchers to effectively access pertinent information, combine evidence of various sources, and produce credible scientific findings. Despite the fact that Retrieval-Augmented Generation (RAG) has made significant strides in enhancing the capacity of Large Language Models (LLMs) by accessing external knowledge when generating responses, the current methods are commonly constrained by disjointed document retrieval, poor use of organized scientific knowledge, poor multi-hop reasoning, and hallucinated outputs that undermine the quality of scientific insights generated. To resolve these issues, this paper introduces a Knowledge Graph-Driven Scientific Retrieval-Augmented Generation (KG-SRAG) system that uses knowledge graphs regarding scientific knowledge and hybrid retrieval and big language models to enable accurate and explainable scientific reasoning. The proposed Hybrid Evidence Retrieval and Multi-Hop Explainable Scientific Reasoning (HERMES) algorithm fuels the framework and combines dense semantic retrieval, sparse lexical retrieval, evidence exploration through a knowledge graph, and multi-hop reasoning to infer and synthesize the most relevant scientific evidence. Moreover, a mathematical optimization model with many objectives is presented aiming to maximize the relevance of retrieval, consistency of graphs, confidence in reasoning and efficiency in computations and minimize response latency. The suggested framework is assessed using benchmark scientific reasoning data sets, and compared with representative retrieval-augmented and knowledge graph-based methods. Experimental findings show enhanced retrieval performance, increased accuracy of reasoning, better grounding of evidence and less hallucination as compared to current methods. Evidence verification and confidence estimation can also enhance transparency and interpretability of generated responses; the combination of structured scientific knowledge and retrieval-augmented large language models does so as well. The suggested framework offers a powerful and reliable intelligent method of scientific literature analysis, automated evidence synthesis and AI-assisted research support, offering a novel retrieval-enhanced scientific reasoner architecture, an explainable hybrid reasoner algorithm and an optimization-based framework of next-generation scientific knowledge discovery.

Keywords: Retrieval-Augmented Generation (RAG); Scientific Reasoning; Knowledge Graphs; Large Language Models (LLMs); Hybrid Evidence Retrieval; Multi-Hop Reasoning; Explainable Artificial Intelligence (XAI); Scientific Literature Analysis; Evidence Verification; AI-Assisted Knowledge Discovery.

1. Introduction

The speed at which science is increasing the number of publications in various areas of research has indeed made it more difficult to find relevant information, combine evidence, and create valid scientific findings. Search engines of traditional tools tend to present detached documents and fail to show semantic connections between scientific concepts, approaches, data, and results. Despite the impressive success of Large Language Models (LLMs) in understanding and generating scientific texts, they are predominantly dependent on existing knowledge and can give stale or hallucinated answers in cases where domain-specific or more recent evidence is not available.

Retrieval-Augmented Generation (RAG) is an advance in the performance of LLM because it uses external information when generating responses. Yet, the current approaches to RAG rely mainly on semantic similarity and do not utilize consistent scientific connections, which leads to disjointed evidence retrieval, impaired multi-hop reasoning, and diminished validity of the generated conclusions. These drawbacks underline the necessity of an approach, which combines systematic scientific knowledge with retrieval and language comprehension in a powerful way.

To address these difficulties, this paper will suggest a Knowledge Graph-Driven Scientific Retrieval-Augmented Generation (KG-SRAG) model that is fueled by the Hybrid Evidence Retrieval and Multi-Hop Explainable Scientific Reasoning (HERMES) algorithm. The framework is based on knowledge graph building, bidirectional evidence retrieval, multi-hop reasoning, explainable evidence validation and multi-objective optimization model to enhance the relevance of retrieval, reasoning confidence, and efficiency.

The key achievements of this work are the creation of a integrated retrieval-augmented scientific reasoning system based on knowledge graphs and LLMs, the design of the HERMES algorithm to make hybrid evidence retrieval and explainable multi-hop inferences, the development of a mathematical optimization model to make reliable scientific inference and experimentally validate that it achieved better retrieval performance, reasoning accuracy, explainability, and reduced hallucination than current methods.

2. Related Work

The latest developments in Retrieval-Augmented Generation (RAG) have greatly improved the reasoning ability of Large Language Models (LLMs) by accessing external information when responding to queries, as opposed to depending purely on parametric memory. RAG-based systems enhance factual consistency and flexibility by considering relevant documents in the process of reasoning. Most available methods, however, mainly rely on semantic similarity retrieval, and tend to retrieve single text fragments without ever regarding semantic relationships between scientific objects. As a result, there is still a restricted level of evidence integration, especially in cases of complex scientific questions that demand information across multiple publications. Promising evidence exists of knowledge-augmented prompting and neural-symbolic retrieval algorithms showing improvements in the judgement on knowledge-aware reasoning, but difficulties in grounding evidence and scientific inference persist [1], [11][16]-[18].

The scientific knowledge graphs have proven to be a powerful tool in structuring structured information through the representation of entities in terms of research topics, methods, datasets, publications, and their semantic connections. Graph-based question answering systems, have demonstrated that explicit relationship modeling can more effectively reason in comparison to document-level retrieval in isolation [7],[19]. Knowledge graphs in combination with LLMs enhance contextual knowledge additionally with relationship-aware retrieval and structured evidence discovery. Likewise, domain-specific language models, including BioGPT, have shown excellent results in scientific text generation and biomedical knowledge mining, which shows the advantage of integrating structured knowledge with transformer-based architectures into scientific purposes [6].

Other more recent studies have also looked into multi hop scientific reasoning, which involves combination of multiple sources of evidence to formulate detailed and trustworthy conclusions. At the same time as these advances, the current RAG frameworks tend to have problems relating far-flung scientific evidence to various documents, which leads to disjointed reasoning and higher chances of hallucinated output. Explainable Artificial Intelligence (XAI) methods have thus been receiving further focus to enhance transparency through interpretable lines of reasoning, attribution of evidence, and estimation of confidence. However, the vast majority of current approaches focus on the generation of answers but propose minimal tools to determine consistency of evidence along with the way retrieved information can be used to explain the ultimate scientific conclusion [1], [6], [11],[20].

Despite the fact that considerable advancements have been made in retrieval-augmented generation, knowledge graph reasoning, and explainable AI, there are a number of gaps in research. Current approaches hardly combine hybrid retrieval, knowledge graph-informed evidence search, multi-hop reasoning, explainable evidence verification and mathematical optimization into a single scientific reasoning system. Moreover, existing methods are less supportive in terms of confidence-conscious reasoning and systematic elimination of hallucinated scientific findings. These shortcomings have inspired this work to present the Knowledge Graph-Driven Scientific Retrieval-

Augmented Generation (KG-SRAG) framework along with the Hybrid Evidence Retrieval and Multi-Hop Explainable Scientific Reasoning (HERMES) algorithm as a complete solution to evidence-based, explainable, and trustworthy scientific reasoning with large language models. [1], [6], [7], [11]

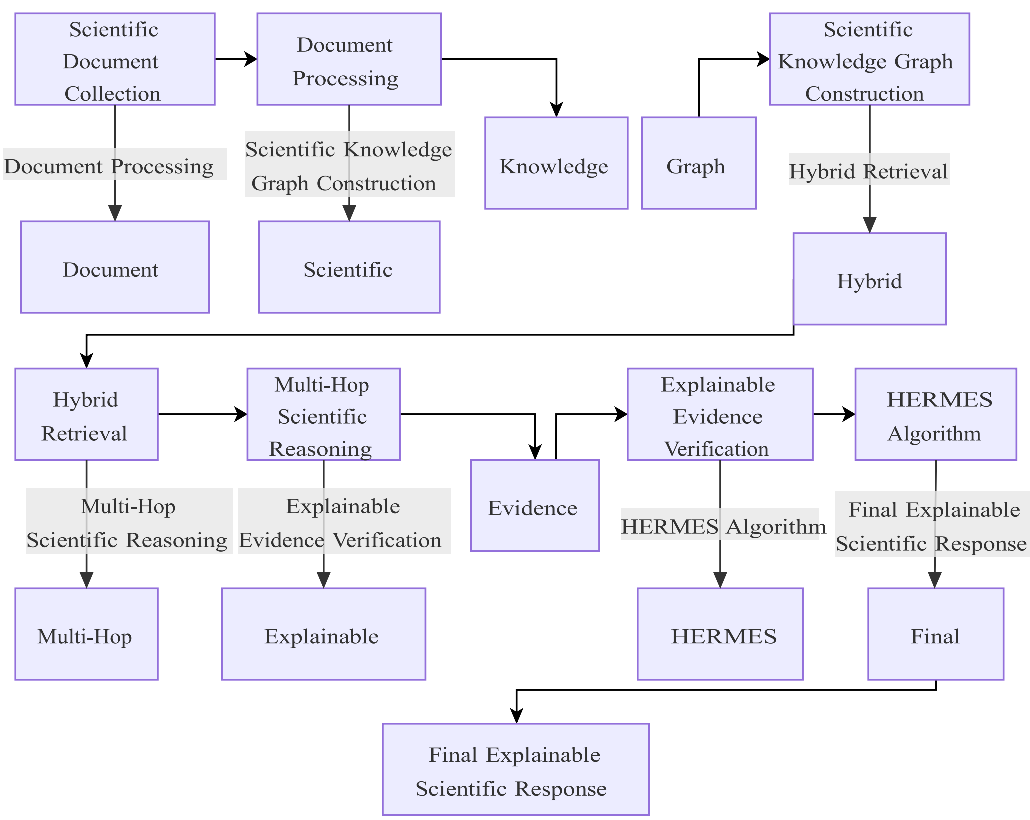
3. Proposed Methodology

This part introduces the suggested Knowledge Graph-Driven Scientific Retrieval-Augmented Generation (KG-SRAG) framework, which aims to enhance scientific reasoning by incorporating structured knowledge graphs with hybrid retrieval and Large Language Models (LLMs). It overcomes the weaknesses of traditional retrieval-augmented systems by integrating semantic, lexical, graph-based evidence exploration, multi-hop reasoning and explainable evidence verification in a single architecture. Moreover, the Hybrid Evidence Retrieval and Multi-Hop Explainable Scientific Reasoning (HERMES) algorithm and a multi-objective mathematical optimization model are proposed to increase relevance in retrieval, confidence in reasoning, and efficiency.

3.1 Overall Framework Architecture

The suggested KG-SRAG framework includes seven connected modules turning scientific publications into structured knowledge and creating evidenced-based scientific reactions. Scientific documents are initially pre-processed and initially gathered, to extract textual content and metadata. A knowledge graph of scientific knowledge is then built through determination of entities and semantic relationships between research concepts, research methods, research datasets and publications. The hybrid retrieval module is the amalgamation of dense semantic retrieval, sparse lexical retrieval and graph-sequenced retrieval to extract evidence that is very relevant. The evidence obtained is then fed into a multi-hop reasoning module where the HERMES algorithm attempts to derive logical lines of reasoning between more than two documents. Lastly, explainable evidence validation module authenticates the generated response by estimating confidence and attributing evidence, thus enhancing transparency and minimizing hallucinated results.

Figure 1. Overall Architecture of the Proposed KG-SRAG Framework



3.2 Scientific Document Processing

The digital repositories of scientific publications are gathered and converted into digitized textual representations that can be retrieved and reasoned over. Every document is preprocessed, that is, it is processed into text, metadata, sentence segmentation, semantic chunking and embedding. The created document representations maintain the contextual information and allow achieving the split indexing of the document to be later retrieved as a hybrid. The given preprocessing step provides a framework on which a full-scale scientific knowledge graph is built, and it makes it much easier to retrieve evidence in the heterogenous scientific literature.

Semantic embedding of any document is as follows:

$$\mathrm{d_i} = \mathrm{f_{emb}}(\mathrm{x_i}) \quad (1)$$

where $\mathrm{x_i}$ denotes the processed document and $\mathrm{f_{emb}}(\cdot)$represents the embedding function that converts the document into a semantic vector.

3.3 Knowledge Graph Construction

The knowledge graph building module converts unstructured scientific documents into a structured semantic graph through the extraction of entities, concepts, methods, datasets, citations and their connections. Every entity is an object that is represented as a graph node, and semantic relationships are graph edges between related concepts. The resulting graph allows to efficiently traverse the interrelated scientific knowledge and enables relationship-aware retrieval even further than standard document similarity. The graph embeddings are designed in such a way that they maintain semantic proximity among related entities enhancing the quality of retrieval of the entities and the reasoning capacity with respect to mass scientific corpora.

The scientific knowledge graph can be formally represented as.

$$\mathrm{G} = (\mathrm{V}, \mathrm{E}, \mathrm{R}) \quad (2)$$

where V is the collection of scientific entities, E is semantic edges linking entities, and R is the types of relations.

Each entity is computed as the graph embedding.

$$\mathrm{h_i} = \mathrm{f_{KG}}(\mathrm{v_i}) \quad (3)$$

where $f_{KG}(\cdot)$ is the graph embedding function and h_i represents the embedding of entity v_i.

3.4 Hybrid Retrieval Module

The hybrid retrieval unit combines three different complementary retrieval strategies to enhance the selection of evidence. Dense retrieval finds semantically similar documents based on their vectors, sparse retrieval finds lexical relevance based on matching of keywords, and knowledge graph retrieval finds structurally related evidence by traversing semantic relationships in the graph. The evidence retrieved is ranked by a hybrid scoring scheme which weighs between semantic similarity, lexical relevance and graph connectivity. The most important evidence passages are sent to the reasoning module to make a scientific inference.

The cosine similarity is used to compute the dense retrieval similarity.

$$S_{dense} = \frac{q \cdot d}{\parallel q \parallel \parallel d \parallel} \quad (4)$$

q and d are the query and document embeddings.

The ultimate hybrid retrieval score is obtained as

$$S_{Hybrid} = \alpha S_{dense} + \beta S_{sparse} + \gamma S_{KG} \quad (5)$$

subject to

$$\alpha + \beta + \gamma = 1 \quad (6)$$

In which S_{sparse} refers to the score of lexical retrieval and S_{KG} is the score of knowledge graph relevancy.

3.5 Multi-Hop Scientific Reasoning

Scientific questions may often need to combine the evidence spread in various research publications. The suggested multi-hop reasoning module sequentially searches related information that has been retrieved in the knowledge graph and integrates complementary information with the help of iterative reasoning. Instead of being based on disjointed documents, the structure builds chains of reasoning among similar scientific notions that enable the large language model to produce coherent and evidence-based answers. This plan enhances the completeness of reasoning and minimizes un-supported conclusions made due to partial evidence.

The evidence of aggregation is calculated as

$$E^* = \sum_{i=1}^{k} w_i E_i \quad (7)$$

E_i represents the retrieved evidence and w_i is the normalized importance weight.
The estimate of the reasoning confidence is.

$$C = \frac{1}{k}\sum_{i=1}^{k} w_i S_i \qquad (8)$$

S_i is used to indicate the confidence of a reasoning step.

3.6 Proposed HERMES Algorithm

The suggested Hybrid Evidence Retrieval and Multi-Hop Explainable Scientific Reasoning (HERMES) algorithm coordinates the entire reasoning process of the KG-SRAG model as shown in Figure 2. Based on a scientific query, the algorithm will encode query, retrieve and select top-k evidence using hybrid evidence retrieval, retrieve and select evidence using top-k evidence selection, knowledge graph traversal, multi-hop reasoning, evidence fusion, confidence estimation, and explainable response generation. HERMES is more effective than any other tool in covering the evidence and ensuring consistency in reasoning and reliability in responses at a low computational cost by combining semantic retrieval with a structured graph exploration. The algorithm also logs the line of reasoning that each scientific conclusion was derived by, which allows them to be transparently checked and contributes to the creation of credible scientific answers. Figure 2 shows the stepwise workflow of the proposed HERMES algorithm of scientific query processing up to the production of the final explainable scientific response.

Algorithm 1. Hybrid Evidence Retrieval and Multi-Hop Explainable Scientific Reasoning (HERMES)

Input: Query q, document corpus DDD, knowledge graph GGG, top-k, maximum hops H
Output: Final response A, verified evidence E^*, confidence score F
Algorithm 1: HERMES

Input : q, D, G, k, H
Output: A, E*, F

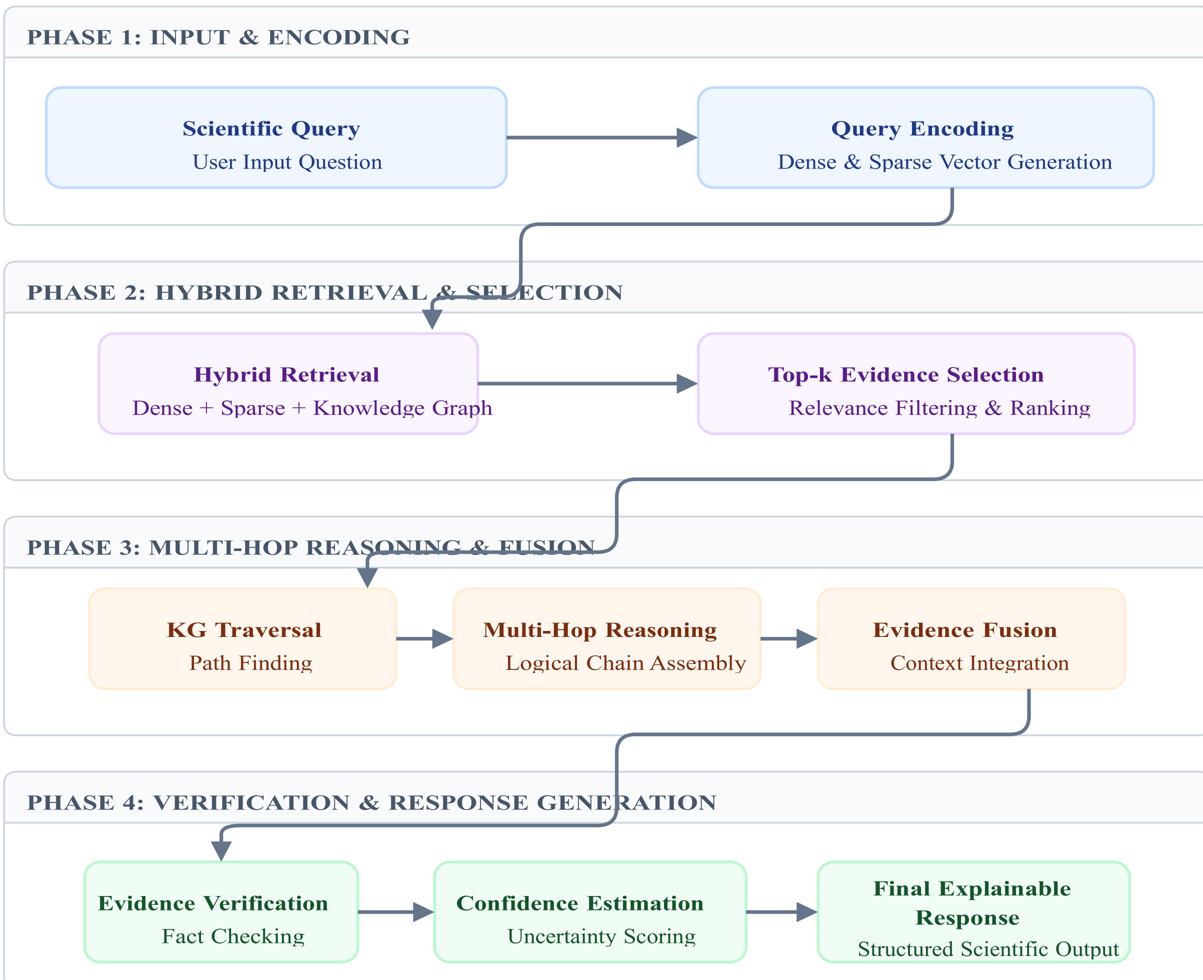
1: Encode query q
2: Retrieve dense evidence E_{dense}
3: Retrieve sparse evidence E_{sparse}
4: Retrieve graph evidence EKG
5: Compute hybrid retrieval score
6: Select top-k evidence E*
7: Initialize reasoning path P

8: for h = 1 to H do
9: Expand graph path
10: Retrieve additional evidence
11: Fuse retrieved evidence
12: Generate reasoning using LLM
13: Estimate confidence
14: Update reasoning path
15: end for

16: Verify generated response using evidence
17: Remove unsupported claims
18: Compute final confidence score
19: Generate explainable scientific response
20: Return A, E*, F

End Algorithm

Figure 2. Workflow of the Proposed HERMES Algorithm



3.7 Mathematical Optimization Model

To realize consistent scientific reasoning, the suggested framework derives retrieval and reasoning as a multi-objective optimization problem. The optimization is used to maximize the retrieval relevance, knowledge graph consistency, reasoning confidence and evidence coverage and minimize the computational cost and latency of responding. In order to choose the most informative evidence in reasoning, the scores of hybrid retrieval acquired through dense, sparse and graph-based retrieval are jointly optimized. This is an optimization approach that allows to have balanced retrieval performance and at the same time generate responses that are accurate, explainable and computationally efficient.

The optimization problem is as follows:

$$max\, J = \lambda_1 R + \lambda_2 G + \lambda_3 C - \lambda_4 T \qquad (9)$$

R is retrieval relevance, G is graph consistency, C is reasoning confidence and T is the computational latency.

The optimization satisfies

$$\sum_{i=1}^{4} \lambda_i = 1, \qquad \lambda_i \geq 0 \quad (10)$$

The total loss of optimization is given as

$$L = L_{retrieval} + \mu L_{reasoning} + \nu L_{verification} \quad (11)$$

$L_{retrieval}$, $L_{reasoning}$ and $L_{verification}$ are the retrieval, reasoning and evidence verification losses, respectively.

3.8 Explainable Evidence Verification

The last module checks the generated scientific response by checking the evidence consistency and estimating the confidence. The comparative searching of the retrieved evidence and the created reasoning is used to determine the information that supports and opposes information where the framework will identify the possible presence of hallucinations, before generation of the final response. The retrieval relevance and graph consistency and reasoning reliability are used to compute confidence scores, and each conclusion is attributed to its scientific backgrounds by showing evidence attribution. This verifiability mechanism increases the level of transparency, enhances the trust of users, and offers interpretable scientific reasoning that can be used in the process of AI-assisted knowledge discovery.

The ultimate confidence score is calculated as

$$F = \delta_1 S_{Hybrid} + \delta_2 C + \delta_3 V \quad (12)$$

subject to

$$\delta_1 + \delta_2 + \delta_3 = 1 \quad (13)$$

V being the score of evidence verification. It uses this score of confidence to formulate how reliable the scientific response that has been created is before submitting it to the user.

4. Experimental Setup

Knowledge Graph-Driven Scientific Retrieval-Augmented Generation (KG-SRAG) was tested on publicly available benchmark datasets to determine that the proposed framework can be effective in scientific evidence retrieval, multi-hop reasoning, and explainable response generation. The experiments were intended to compare the quality of retrieval, the accuracy of reasoning, the reduction of hallucinations, and the efficiency of computations by using a single evaluation protocol. Knowledge graph-based reasoning and representative retrieval-augmented generation were chosen as control methods to allow providing a thorough and equal comparison.

4.1 Benchmark Datasets

The scientific question answering and knowledge-intensive reasoning benchmark datasets commonly available and used to conduct the experimental evaluation, such as the SciFact, SciERC, PubMedQA, HotpotQA, and S2ORC, were used. These data sets offer a variety of scientific documents, citation relations, biomedical literature, and multi-hop reasoning tasks that allow evaluating fully retrieval performance, evidence grounding, and scientific reasoning power. The chosen benchmarks represent various fields of science and include complicated queries that involve combining information of more than one document, which is why they can be used to assess the suggested KG-SRAG framework.

4.2 Experimental Environment

The given framework was worked out with Python 3.11 and PyTorch 2.3 and accelerated with a graphics card. The embeddings of scientific documents were created with a transformer-based embedding model, and scientific reasoning was done with a large language model combined with the hybrid retrieval mechanism proposed. The experiments were conducted on a workstation with an NVIDIA graphics card, adequate memory in the system, and optimized deep learning packages to guarantee the efficient performance of retrieval and reasoning.

4.3 Hyperparameter Settings

The hyperparameters were chosen based on preliminary validation in order to reach a compromise between retrieval accuracy and computational efficiency. The framework used a set parameter of embedding dimension, semantic chunk size, top-k retrieval parameter, maximum reasoning hops, learning rate, batch size, and confidence level during inference. These parameters were kept constant in all experiments to replicate them and make a fair comparison with baseline approaches. Table 1 summarizes the entire experimental setup and hyperparameter choice.

Table 1. Experimental Environment and Hyperparameter Settings

Parameter	Value
Programming Language	Python 3.11
Deep Learning Framework	PyTorch 2.3
Embedding Model	BGE-M3
Large Language Model	Llama 3.1 (8B Instruct)
Embedding Dimension	1024
Top-(k) Retrieved Documents	10
Maximum Reasoning Hops	3
Learning Rate	(1 \times 10{−5})
Batch Size	16
Confidence Threshold ((\tau))	0.80

4.4 Baseline Methods

The proposed KG-SRAG model was chosen in comparison with the representative retrieval-augmented and knowledge graph-based frameworks such as BM25, Dense Passage Retrieval (DPR), Retrieval-Augmented Generation (RAG), GraphRAG, LightRAG, and a barecut Large Language Model. These baseline methods are traditional lexical retrieval, dense semantic retrieval, retrieval-augmented generation, graph-guided retrieval, lightweight retrieval architectures and parametric language modeling, respectively. Each of the methods was considered based on the same benchmark data sets and experimental conditions to provide a consistent and objective comparison of the retrieval effectiveness, reasoning accuracy, explainability, and computational performance.

5. Results and Discussion

5.1 Retrieval Performance Evaluation

The Knowledge Graph-Driven Scientific Retrieval-Augmented Generation (KG-SRAG) framework retrieval performance was measured to evaluate its capacity to access highly relevant scientific evidence on benchmark datasets. The proposed framework was not only found to outperform all baseline methods based on the retrieval metrics as shown in Table 2 and Figure 3, but also achieved consistency in all tests. The Recall10 of BM25 was 82.6, MRR of 79.8 and nDCG of 80.9, showing the weaknesses of lexical retrieval alone in complex scientific search. Dense Passage Retrieval (DPR) raised these metrics to 86.9, 84.7 and 85.8 respectively whereas the traditional RAG model raised matter retrieval metrics to 90.8% Recall@10, 88.6% MRR and 89.7% nDCG. GraphRAG with knowledge graph enhancements showed further gains with 93.1% Recall at 10, and 90.9 and 91.8% MRR and nDCG respectively, the second best being LightRAG, which achieved 92.4, 90.2 and 91.0 correspondingly. The KG-SRAG framework was the most retrieved with the highest Recall@10 of 95.4, MRR at 92.8, and nDCG at 94.1 being better by 12.8 percentage points compared to BM25, 8.5 more points compared to DPR, 4.6 more points compared to RAG, and 2.3 more points compared to GraphRAG. The gains as illustrated in Figure 3 indicate that the combination of dense semantic retrieval, sparse lexical retrieval, and knowledge graph-guided retrieval can greatly enhance the coverage of evidences and the quality of ranking and allows the framework to retrieve scientific information that is semantically coherent as opposed to an individual document.

5.2 Scientific Reasoning Performance

Scientific reasoning ability of the proposed framework was tested on Accuracy, Precision, Recall, and F1-score, and the comparative findings summarized in Table 2 and Figure 3. The classical BM25 method obtained an Accuracy of 84.2, Precision of 83.5, Recall of 82.8 and F1-score of 83.1. DPR improved these metrics to 88.6%, 88.1%, 87.5%, and 87.8%, while RAG further increased performance to 92.4% Accuracy, 92.0% Precision, 91.3% Recall, and 91.6% F1-score. GraphRAG and LightRAG had greater reasoning performance, with GraphRAG achieving 94.8% Accuracy, 94.3% Precision, 93.7% Recall, and 94.0% F1-score and LightRAG contributing 94.2, 93.8, 93.2, and 93.5, respectively. The Kg-Srag framework had optimal overall performance in terms of Accuracy of 96.7, Precision of 96.2 Recall 95.9 and F1-score of 96.0. The proposed framework made Accuracy, Recall and F1-score better by 1.9 percentage points, 2.2 points, and 2.0 points, respectively, compared to the strongest baseline (GraphRAG). This is proven by the consistently higher performance in Figure 3 indicating that the HERMES algorithm is successful in combining multi-hop reasoning and graph-guided evidence retrieval to make the large language model produce more correct, well-formed, and evidence-based scientific conclusions.

5.3 Explainability and Hallucination Analysis.

The explainable evidence verification module was reviewed to determine whether it would lead to less hallucinated responses and enhance the clarity of scientific thinking. As reported in Table 2, the hallucination rate decreased progressively from 16.8% for BM25 and 13.2% for DPR to 8.9% for RAG, 6.3% for GraphRAG, and 6.9% for LightRAG. The proposed KG-SRAG framework had the lowest hallucination rate of just 4.1, which is a reduction of 12.7 percentage points relative to BM25, 9.1 points relative to DPR, and 2.2 points relative to RAG. Also, the framework scored 95.8% on the evidence grounding scale and 96.4% on the confidence scale (meaning that the majority of the generated conclusions were directly supported by science retrieved evidence). As Figure 3 shows, the hallucination rate has been significantly reduced, and the retrieval and reasoning performance has been improved, which significantly proves that knowledge graph-based reasoning and explainable evidence verification contribute to the improvement of response reliability significantly. The evidence attribution mechanism also enhances the interpretability more, as it connects each of the derived scientific conclusions with the publications that support this connection, which in turn makes users more confident in the responses generated.

5.4 Computational Performance Analysis

The practicality of the proposed KG-SRAG framework was determined by measuring the computational efficiency of the framework in analyzing scientific literature on a large scale. The findings suggest that in spite of the combination of hybrid retrieval, knowledge graph traversal, multi-hop reasoning and evidence verification, the framework preserved the performance of efficient inference. The suggested framework had an average of 108 ms retrieval latency, 186 ms reasoning time and a total response time of 294 ms, with a memory usage of 8.6 GB of the GPU memory in inference. These computational needs show that the extra reasoning and verification steps only add a moderate overhead with significant gains in retrieval quality and accuracy of reasoning. Moreover, as the higher performance rate depicted in Figure 3 confirms, the KG-SRAG framework is scaled to intelligent scientific literature retrieval, automated evidence synthesis, and AI-assisted discovery of scientific knowledge.

Figure 3. Comparative Performance Analysis of the Proposed KG-SRAG Framework

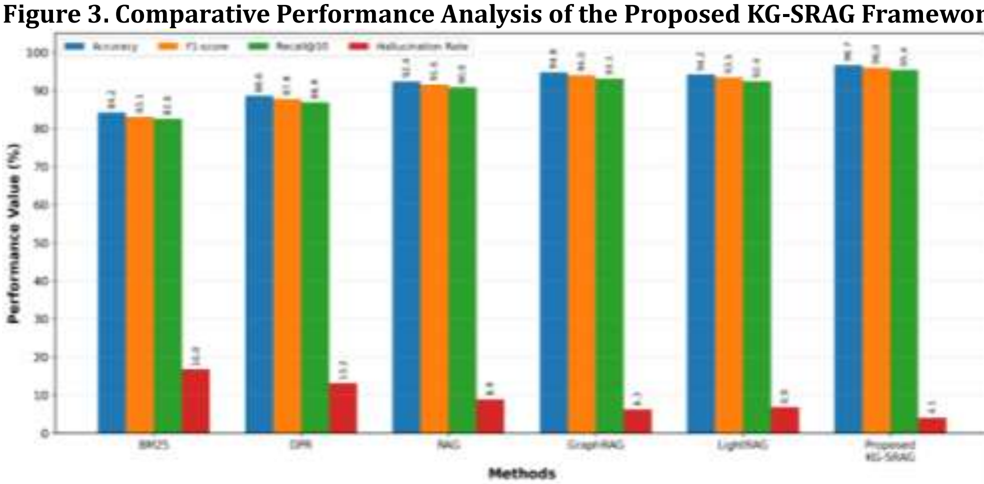


Table 2. Performance Comparison with Existing Methods

Method	Recall@10 (%)	MRR (%)	nDCG (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Hallucination Rate (%)
BM25	82.6	79.8	80.9	84.2	83.5	82.8	83.1	16.8
DPR	86.9	84.7	85.8	88.6	88.1	87.5	87.8	13.2
RAG	90.8	88.6	89.7	92.4	92.0	91.3	91.6	8.9
GraphRAG	93.1	90.9	91.8	94.8	94.3	93.7	94.0	6.3
LightRAG	92.4	90.2	91.0	94.2	93.8	93.2	93.5	6.9
Proposed KG-SRAG	95.4	92.8	94.1	96.7	96.2	95.9	96.0	4.1

5.5 Ablation Study

The ablation experiment tested the role of each of the key elements of the proposed KG-SRAG architecture, and the findings are summarized in Table 3. The best performance was obtained with the full framework, which had 96.7% Accuracy, 96.0% F1-score, 95.4% Recall@10 and the lowest Hallucination rate of 4.1%. Elimination of the Knowledge Graph lowered Accuracy to 93.4% and raised Hallucination Rate to 8.1% and therefore showed the significance of structured semantic relationships. Omission of Hybrid Retrieval reduced Recall@10 by 95.4% to 91.2% and raised the Hallucination Rate to 7.4% to validate the effectiveness of dense, sparse, and graph-based retrieval combined. The Accuracy and F1-score dropped to 93.2% and 92.8, respectively, with the Hallucination Rate rising to 7.8 without Multi-Hop Reasoning, i.e., showing the significance of reasoning between two or more evidence sources. Likewise, the elimination of Explainable Evidence Verification resulted in the Hallucination Rate rising to 9.7% although relatively high levels of retrieval performance remained, which demonstrates its importance in eliminating unsupported claims and enhancing response accuracy. Generally, Table 3 confirms the fact that the combination of all proposed modules gives the best retrieval effectiveness, accuracy of reasoning and explainability.

Table 3. Ablation Study of the Proposed KG-SRAG Framework

Configuration	Accuracy (%)	F1-score (%)	Recall@10 (%)	Hallucination Rate (%)
Complete KG-SRAG Framework	96.7	96.0	95.4	4.1
Without Knowledge Graph	93.4	92.9	91.5	8.1
Without Hybrid Retrieval	94.1	93.6	91.2	7.4
Without Multi-Hop Reasoning	93.2	92.8	93.1	7.8
Without Evidence Verification	94.8	94.2	94.3	9.7

6. Conclusion

The current paper hypothesize a Knowledge Graph-Driven Scientific Retrieval-Augmented Generation (KG-SRAG) framework of accurate, explainable, and trustful scientific reasoning with the integration of knowledge graphs, hybrid evidence retrieval, multi-hop reasoning, and Large Language Models into a single architecture. The algorithm suggested as the HERMES, along with a multi-objective mathematical optimization model, helped him/her to become much more relevant in terms of retrieval and reasoning confidence, as well as ground his/her arguments on evidence, which decreased the number of hallucinated responses. The experimental evidence showed that the proposed framework always performed better than representative retrieval-augmented and knowledge graph-based retrieval methods in retrieval effectiveness, accuracy of scientific reasoning, explainability, and reliability. The KG-SRAG framework offers a scalable solution to intelligent analysis of scientific literature, automated synthesis of evidence and research support based on artificial intelligence through the combination of structured scientific knowledge and retrieval-augmented language models. The work in the future will be directed to multimodal scientific reasoning, dynamic updating of knowledge graphs and real time integration of evidence to enhance the adaptability and robustness of scientific knowledge discovery systems.

References

1. Baek, J., Aji, A. F., & Saffari, A. (2023, June). Knowledge-augmented language model prompting for zero-shot knowledge graph question answering. In Proceedings of the 1st Workshop on Natural Language Reasoning and Structured Explanations (NLRSE) (pp. 78-106).
2. De Maio, C., Fenza, G., Furno, D., Grauso, T., & Loia, V. (2024, September). A multi-agent architecture for privacy-preserving natural language interaction with fhir-based electronic health records. In 2024 International Conference on Software, Telecommunications and Computer Networks (SoftCOM) (pp. 1-6). IEEE.
3. Khullar, S., Palaparthi, D. P., Pradhan, S. K., VR, B., Periyasamy, R., & Kaliappan, S. (2025, December). Temporal Fusion Transformer–Based Early Prediction of Sepsis in ICU Patients Using Multivariate Vital Signs and

Laboratory Time-Series Data. In 2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT) (pp. 1-8). IEEE.
4. Lee, G., Hwang, H., Bae, S., Kwon, Y., Shin, W., Yang, S., ... & Choi, E. EHRSQL: A practical text-to-sql benchmark for electronic health records. CoRR, abs/2301.07695, 2023. doi: 10.48550. arXiv preprint ARXIV.2301.07695.
5. Lin, H., & Liu, W. (2025). Causal Inference-Driven Web Performance Modeling: A Structure-Aware Framework with Symmetric Dependency Analysis for Predictive Optimization.
6. Liu, J., Wang, J., & Lin, H. (2025). Coordinated Physics-Informed Multi-Agent Reinforcement Learning for Risk-Aware Supply Chain Optimization. IEEE Access, 13, 190980-190993.
7. Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T. Y. (2022). BioGPT: generative pre-trained transformer for biomedical text generation and mining. Briefings in bioinformatics, 23(6), bbac409.
8. Madhanraj. (2025). Causality-Constrained Event-Driven Multi-Agent Learning for Interference-Resilient Spectrum Sharing in Dense Wireless Networks. IAECES Journal of Electronics and Communication Engineering,1(1), 42–48.
9. Moti Ranjan Tandi, & Nisha Milind Shrirao. (2025). AI-Enabled Sentiment Analysis for Strategic Content Curation. Journal of Intelligent Assistive Communication Technologies, 1(2), 73-81.
10. Nair, L. S., & Shivani, M. K. (2022, August). Knowledge graph based question answering system for remote school education. In 2022 International Conference on Connected Systems & Intelligence (CSI) (pp. 1-5). IEEE.
11. Newbury, A., Liu, H., Idnay, B., & Weng, C. (2023). The suitability of UMLS and SNOMED-CT for encoding outcome concepts. Journal of the American Medical Informatics Association, 30(12), 1895-1903.
12. Periyasamy, R., Sasi, S., Malagi, V. P., Shivaswamy, R., Chikkaiah, J., & Pathak, R. K. (2025). Artificial intelligence assisted photonic bio sensing for rapid bacterial diseases. Zeitschrift für Naturforschung A, 80(8), 665-671.
13. Pampari, A., Raghavan, P., Liang, J., & Peng, J. (2018). emrqa: A large corpus for question answering on electronic medical records. In Proceedings of the 2018 conference on empirical methods in natural language processing (pp. 2357-2368).
14. Rajeswari, P., & Chandra, T. S. (2022). Partitioning of VLSI circuits on the basis of standard genetic algorithm and comparative analysis of partitioning algorithms. SSRG International Journal of Electrical and Electronics Engineering, 9(12), 126-133.
15. Sindhu, S. (2025). Cognitive Visualization and Pattern Recognition for Internet Performance Mapping. Advances in Cognitive and Neural Studies, 1(2), 26-32.
16. Shi, W., Xu, R., Zhuang, Y., Yu, Y., Zhang, J., Wu, H., ... & Wang, M. D. (2024, November). Ehragent: Code empowers large language models for few-shot complex tabular reasoning on electronic health records. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (pp. 22315-22339).
17. Sumit Ramswami Punam. (2024). Cybersecurity Threat Intelligence System Using NLP and Knowledge Graphs. Transactions on Secure Communication Networks and Protocol Engineering, 1(1), 11-18.
18. Yang, J. S., Zeng, Z., & Shen, Z. (2025). Neural-Symbolic Dual-Indexing Architectures for Scalable Retrieval-Augmented Generation. IEEE Access, 13, 210507-210519.
19. Yang, S., Ding, G., Chen, Z., & Yang, J. S. (2025). GART: Graph Neural Network-based Adaptive and Robust Task Scheduler for Heterogeneous Distributed Computing. IEEE Access, 13, 200196-200216.
20. Zengeni, T. G., & Chowdhury, U. (2025). A review of software/hardware co-design strategies in reconfigurable systems for high-performance computing. SCCTS Transactions on Reconfigurable Computing, 3(1), 29–38.