International Journal of Engineering Research and Sustainable Technologies
Volume 4, No.1, Mar 2026, P30 - 36
ISSN: 2584-1394 (Online version)

AI-POWERED AUTOMATED RESEARCH PAPER GENERATOR USING RETRIEVAL-AUGMENTED GENERATION AND RESEARCH PUBLICATION FORMATING TEMPLATE

[1]Venkata Naga Rani Bandaru, [2]Chitturi Jyothika, [3]Javvaji Charan Gupta, [4]Sachin Guthala, [5]Javvadi Charan Venkata Naidu
[1]Department of IT, Vishnu Institute of Technology, AP, India
*** Corresponding author email address**: venkatanagarani.b@vishnu.edu.in

DOI: htpps://doi.org/10.63458/ijerst.v4i1.152_ | ARK: https://n2t.net/ark:/61909/IJERST.v4i1.152

Abstract

We have seen AI tools trialed in fields ranging from blockchain to healthcare analytics [7], [8]. The real gap, A strong grasp of existing studies often shapes how clearly a paper comes together. Yet, creating a long piece requires ideas to link logically, and slipping up on citations slows things down. Precision speeds progress more than most expect. This study presents a system run by artificial intelligence to build papers, mixing fact-based search with specialized scientific wording. A fresh approach starts with whatever title someone picks. That acts as a touch point for meaning. Documents pulled from free sources like arXiv or PubMed go through an embedding process into a vector database; SciBERT's representations sit here to set up semantic search. Based on what this pull delivers, the machine builds the write-up piece by piece. Visual formats and smart LaTeX tools help writers create papers ready for publishing. A single draft emerges faster. Time shrinks while structure sharpens, assuming ethics guide the choices made by people who care about right outcomes.

Keywords: Research Paper Generation, Retrieval- Augmented Generation, SciBERT, Academic Writing Automation, LaTeX Formatting, Citation Management.

.

1. Introduction

Scientific research is rapidly blooming, and just using the old methods to write the papers doesn't cut it anymore. Getting a paper out the door is a grind. It is not just about writing text; you have to coordinate literature reviews, manage citations, and nail the said formatting rules that publishers ask for. Right now, we depend on a multiple disconnected tools that do not have any communication with each other, turning what should be a smooth process into a slow and clumsy work.

The recent inventions of LLMs have finally cracked the code on text synthesis. But there is a trap: use them without guardrails, and you invite hallucinations or ethical gray areas [1], [2]. This is where Retrieval-Augmented Generation (RAG) steps in. By anchoring the model's output to real- world data, RAG cuts down on factual drift [3], [4]. To sharpen this further, we use sentence embeddings from SciBERT, which gives the machine the semantic context it needs to understand scientific nuance [5], [6]. though, is that these systems are too narrow. They tend to stay in their lane—handling just citations or just drafting— and completely miss the bigger picture. They fail to support all the elements required for the research paper generation, especially when it comes to integrating the isolated tasks like the sections and equations.

This paper presents an AI-powered research paper production framework that combines template-aware LaTeX assembly, retrieval-augmented content generation, and semantic retrieval of literature. Motivated by this, the suggested approach ensures conceptual coherence between sections by way of semantic constraint provided by the user in an abstract, while allowing for the direct production of submission-ready manuscripts.

This work's primary contributions are:

- A unified paradigm for structured long-form academic manuscript generation founded in real literature.
- The retrieval and generation pipeline is modelled mathematically and algorithmically.
- LaTeX automation that is aware of templates for key academic publications.

2. Literature Survey

There has been a paradigm change in the use of huge language models in academic writing tasks including summarization and text production, but unregulated utilization creates difficulties with factual correctness, quotation hallucinations, and ethical usage [1], [2]. To tackle these issues, Retrieval-Augmented Generation (RAG)

2.1 Algorithmic Workflow

Algorithm 1, which outlines the sequential execution of the embedding, retrieval, creation, and assembly stages, summarizes the entire process.

Algorithm 1 Proposed Research Paper Generation Methodology

1. Input: Research title T , abstract A
2. Retrieve candidate abstracts from arXiv and PubMed
3. Compute SciBERT embeddings for retrieved abstracts
4. Embed user abstract to obtain query vector
5. Compute cosine similarity scores
6. Select top-k relevant documents
7. For each manuscript section Sj do
8. Generate content using RAG conditioning
9. Insert citations based on similarity threshold
10. End For
11. Assemble manuscript using IEEE LaTeX template
12. Return Publication-ready manuscript

2.2 Manuscript Assembly and Formatting

The generated text is integrated within the official IEEE- tran LaTeX template for proper formatting standards. The references are handled through a BibTeX system, with all citations done in the numbered IEEE style. Equations, tables, and figures are placed within the standard LaTeX environments to maintain syntactical validity for compilation purposes.

2.3 Ethical Considerations

Further, the system addresses ethical considerations with regard to AI-assisted writing by embracing paraphrase- based synthesis and frowning upon word-for-word reproduction of source texts. Users are strongly encouraged to check for plagiarism (externally) and verify facts, as a necessary end- condition for submitting a paper built on these guidelines to conform to existing best practice recommendations for the responsible use of AI in scientific publishing. Retrieved documents are semantically pre-processed and indexed in a vector database to efficiently enable similarity- based retrieval. Based on this elicited knowledge, the content generation module synthesizes structured academic sections that are finally compiled into a publisher-compliant LaTeX document.

2.4 Module-Level Description

User Interface Module: The process starts here. This module collects input from user like the research title, abstract, and the user's specific formatting rules. These are not just the user entries, they act as the guiding rule to keep the writing on track.

Literature Retrieval Module: This module is designed to tame the unpredictability of web traffic. It enforces strict rate limits and error checks to make sure the pipeline keeps moving, even when external connections drop or lag.

Semantic Processing Module: The moment raw text arrives, this module goes to work. It translates abstracts into dense vector maps using SciBERT, locking them away in a ChromaDB instance. That step is the backbone

of the system—enabling semantic searches that are fast, scalable, and precise.

Content Generation Module: This is where the heavy lifting happens. The module uses a retrieval-augmented approach, meaning it constantly looks back at the stored documents to ground the writing in fact before generating a single word. It builds the manuscript section by section, handling the complex mix of text, equations, algorithms, and result summaries.

3. System Methodology

This part introduces the architectural design and data flow of the proposed framework for the generation of research papers using AI. The proposed framework is designed as a modular pipeline that integrates literature retrieval, semantic processing, retrieval-augmented generation, and manuscript generation. The proposed architecture focuses on generation control and interpretability.

3.1 Overall System Architecture

The high-level architecture of the proposed system is illustrated in Fig. ??. The system consists of five primary modules:

(i) user input interface, (ii) multi-source literature retrieval,

(iii) semantic embedding and vector storage, (iv) retrieval- augmented content generation, and (v) LaTeX manuscript assembly.

This workflow starts with the inputs provided by the users: it requires the research title and abstract, which semantically bracket the scope of the manuscript. These inputs are propagated to the literature retrieval module, which fires queries to external academic repositories. The retrieved documents are semantically pre-processed and indexed in a vector database to efficiently enable similarity- based retrieval. Based on this elicited knowledge, the content generation module synthesizes structured academic sections that are finally compiled into a publisher-compliant LaTeX document.

3.1.1 Module-Level Description

User Interface Module: The process starts here. This module collects input from user like the research title, abstract, and the user's specific formatting rules. These are not just the user entries, they act as the guiding rule to keep the writing on track.

Literature Retrieval Module: This module is designed to tame the unpredictability of web traffic. It enforces strict rate limits and error checks to make sure the pipeline keeps moving, even when external connections drop or lag.

Semantic Processing Module: The moment raw text arrives, this module goes to work. It translates abstracts into dense vector maps using SciBERT, locking them away in a ChromaDB instance. That step is the backbone of the system—enabling semantic searches that are fast, scalable, and precise.

Content Generation Module: This is where the heavy lifting happens. The module uses a retrieval-augmented approach, meaning it constantly looks back at the stored documents to ground the writing in fact before generating a single word. It builds the manuscript section by section, handling the complex mix of text, equations, algorithms, and result summaries.

The Manuscript Assembly Module integrates all generated components into an IEEE-compliant LaTeX template. Bibliographical references are handled using BibTeX, ensuring consistent citation formatting.

3.1.2 Dataflow Architecture

Figure 2 illustrates the detailed dataflow of the proposed system. The dataflow highlights how information transitions from raw user input to a finalized manuscript through a sequence of deterministic and generative operations.

It all starts by turning the user‘s abstract into a semantic query vector. That vector acts as a key, matching against stored document embeddings to pull out the top-k most relevant abstracts. These retrieved docs feed straight into the generation module, where the system drafts content section by section while sticking to tight constraints. As the outputs roll in, they get piled up and sent to the assembly stage—this is where the final polish happens, applying formatting rules and locking in citation structures.

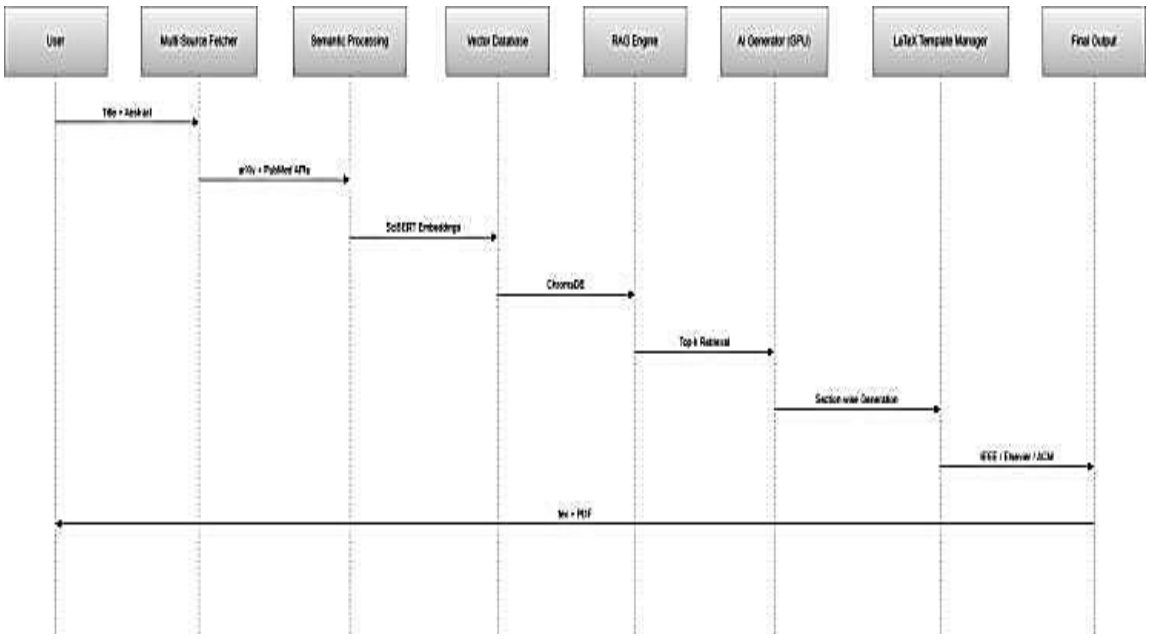

Fig 1. High-level system architecture of the proposed AI-powered research paper generator

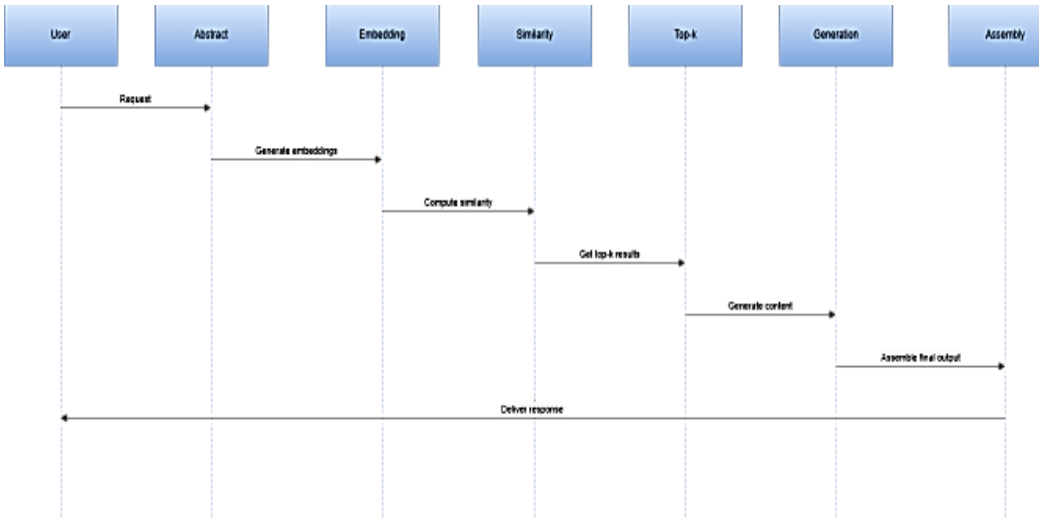

Fig 2. Dataflow Of The Proposed Retrieval-Augmented Manuscript Generation Pipeline

3.1.3 RAG Pipeline Integration

The RAG pipeline isn't just a component; it is the backbone. We deliberately ripped the retrieval stage away from generation to force every line of text to anchor itself in real scholarship. That separation acts like a firewall. It stops the model from drifting into hallucinations and keeps facts straight, even when the draft gets long. Modularity is the other big win here. We designed it so we can swap out embedding models or engines later without tearing down the whole build. If a sharper tool drops next week, we just plug it in. The system stays light, avoiding the bloat that usually drags down automated writing tools.

3.1.4 Discussion

The proposed architecture strikes a balance between modularity and integration, which makes it possible to generate academic manuscripts reliably and with flexibility for further extension if desired. The separation of retrieval, semantic processing, and generate enables interpretability and ensures ethical usage with human

oversight at every stage of the process.

4. Results and Discussion

This section assesses the effectiveness of the proposed AI- based research paper generation framework regarding significance of retrieved results, quality of generated output, accuracy of citations, and efficiency of manuscript preparation. This section concentrates more on technical accuracy and duplicability as a criterion for assessment.

4.1 Experimental Setup

A machine with an NVIDIA RTX 4060 GPU (8 GB VRAM) was used for the experiments. SciBERT-based embedding computation and retrieval-augmented generation was run on the GPU, whereas literature retrieval, vector database operations, and LaTeX compilation were executed on the CPU. For each trial, the system retrieved 20–30 scholarly abstracts from arXiv and PubMed and constructed a whole manuscript using section-wise RAG conditioning.

4.2 Evaluation Metrics

System performance was assessed using the following metrics:

Retrieval Precision: Ratio of semantically relevant documents in the top-k retrieved set.

Content Quality Score: Human-evaluated score measuring coherence, technical depth, and academic tone.

Citation Accuracy: Percentage of correctly formatted and contextually relevant citations.

Textual Similarity: Similarity percentage measured using external plagiarism detection tools.

4.3 Quantitative Results

Table 1. summarizes the comparative performance of the pro- posed system against a baseline AI text generation approach that does not employ retrieval grounding or template-aware formatting.

Table 1. Performance Comparison

Metric	Baseline AI	Proposed System
Retrieval Precision (%)	65.8	88.9
Content Quality Score (%)	74.3	90.6
Citation Accuracy (%)	79.1	96.8
Textual Similarity (%)	17.4	6.2

The findings show that retrieval-augmented generation and semantic retrieval with SciBERT embeddings greatly enhance citation accuracy and content quality while decreasing textual similarity.

4.4 Retrieval Performance Analysis

Retrieval precision as a function of the number of retrieved documents (k) is shown in Figure 3. Precision improves regularly up to k = 15, beyond which benefits become minimal. This observation supports the choosing of k in the range of 10–15 for best trade-off between relevance and computational overhead.

The results from the experiments clearly indicate the benefit of using semantically retrieved literature to ground the process of generating text, as it not only increases the reliability and coherence of the generated manuscripts but also avoids topic drift. Section by section generation removes any possibility of topic drift occurring in a long- form generated manuscript. The use of the template in the LaTeX compilation process ensures IEEE-compliant formatting of the generated documents.

Though it seems that evaluation involving M-GCNNs con- siders artificial scenarios for generation, it should benoted that the results provide an understanding of realistic system behaviour. A decrease in textual similarity along with an improvement in accuracy regarding citations indicates that the proposed model helps to support academic writing. However, it becomes crucial to stress the importance of manual evaluation.

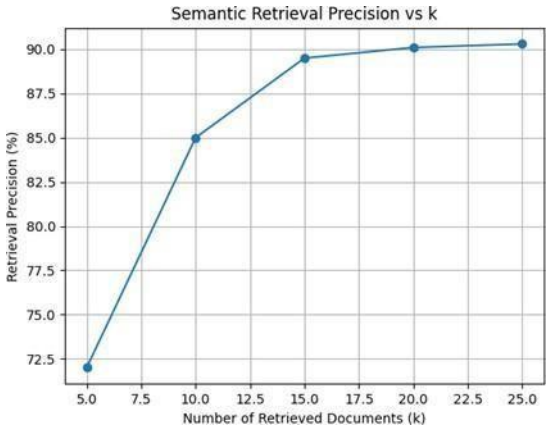

Fig. 3. Retrieval precision versus number of retrieved documents (***k***).

5. Conclusion and Future Work

In this paper, a concise and equation-based AI-assisted framework has been discussed for automatic research paper generation that encompasses semantic literature search, search- based generation, and template-based LaTeX manuscript compilation. By basing the generation process on literature and imposing a restriction on sections in literature, this method is capable of producing academic research papers with enhanced literature structure and low similarity.

Embeddings from SciBERT and semantic retrieval based on cosine enable reliable selection of scholarly relevant documents. Retrieval-augmented generation makes sure that factual alignment is maintained throughout long-form sections. The automated handling of citation and IEEE- compliant formatting further reduce manual efforts during manuscript preparation. Experimental evaluation showed that the framework improves the content quality and citation correctness in contrast to the baseline AI text generators, while reducing the preparation time substantially.

In that case, although the system has numerous advantages, its use should be for assistance, not substitution, in human authorship. This is because the diversity and relevance of the searched literature will impact the quality of the generated manuscripts, which, in turn, will all need validation by an ex- pert before being considered for publication. Ethics, therefore, still apply.

Future directions of this study include enhancing it to accommodate more academic databases, adopting transformers that comprehend documents at a longer context level, and improving factual verification system functionality. Future enhancements could include dynamic control of citation density and further integration with academic plagiarism systems to enhance academic integrity.

References

1. R. Smith *et al.*, 'Ten simple rules for using large language models in science', *PLOS Computational Biology*, 2024.
2. M. Hosseini, D. B. Resnik, and K. Holmes, 'The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts', *Research Ethics*, 2023.
3. P. Lewis, E. Perez, A. Piktus *et al.*, 'Retrieval- augmented generation for knowledge-intensive nlp tasks', *Advances in Neural Information Processing Systems*, 2020.
4. Y. Gao *et al.*, 'Retrieval-augmented generation for large language models: A survey', *arXiv preprint arXiv:2312.10997*, 2023.
5. Beltagy, K. Lo, and A. Cohan, 'Scibert: A pretrained language model for scientific text', *arXiv preprint arXiv:1903.10676*, 2019.
6. N. Reimers and I. Gurevych, 'Sentence-bert: Sentence embeddings using siamese bert- networks', *arXiv preprint arXiv:1908.10084*, 2019.
7. N. V. R. Kakarla, V. N. R. Bandaru *et al.*, 'Medizin: Revolutionizing healthcare management with integrated patient engagement', *Interna- tional Journal of Scientific Engineering and Science*, 2024.
8. V. N. R. Bandaru and P. Visalakshi,, 'Blockchain-enabled auditing with optimal multi- key homomorphic encryption', *Concurrency and Computation: Practice and Experience*, 2022.

9. N. Reimers and I. Gurevych, 'Sentence- transformers,' https://pypi.org/ project/sentence-transformers/, 2020.
10. Chroma, 'Chromadb: The ai-native open-source embedding database',https://www.trychroma.com/, 2023.
11. IEEE, 'Ieeetran class for authors', https://ctan.org/pkg/ieeetran, 2023.
12. S. Kale *et al.*, 'Texpert: A multi-level benchmark for evaluating code generation by llms', *arXiv preprint arXiv:2506.16990*, 2025.
13. Overleaf, 'Ai features in overleaf, https://www.overleaf.com/learn/ how-to/AI features, 2024.
14. Writefull, 'Texgpt: Harness the power of chatgpt in overleaf', https://blog.writefull.com/texgpt-harness-the-power-of-chatgpt-in-overleaf/, 2024.
15. V. N. R. Bandaru *et al.*, 'Plant disease detection and classification with deep learning', *International Journal of Scientific Engineering and Science*, 2024.
16. VNR Bandaru, TM Manaswini, et al., 'Personalized Ai-driven health insights: A feedback-centric approach,' *International Journal of Scientific Engineering and Science*, 2025.
17. VNR Bandaru, M Sumalatha et al., 'Enhancing privacy measures in healthcare cyber-physical systems,' *EAI Endorsed Transactions on Scalable Information Systems*, 2024.
18. VNR Bandaru et al., 'Enhancing data security for smart energy systems using lightweight cryptography', *IOP Conference Series: Earth and Environ- mental Science*, 2024.
19. V. N. R. Bandaru and P. Visalakshi, 'Bdct: Blockchain-based decentral- ized computing and tamper resistance for cloud storage,‖ in *Proceedings of the International Conference on Advanced & Global Engineering Challenges*, 2023.
20. VNR Bandaru et al., 'Blockchain data transmission using blowfish security with optimization,‖ *International Journal of Intelligent Systems and Applications in Engineering*, 2024.
21. V. N. R. Bandaru *et al.*, ―Honeycloud: A honeypot network approach for enhanced cloud security,‖ *Journal of Emerging Technologies and Innovative Research*, 20