



Next-Generation Educational AI Using Retrieval-Augmented Generation: Research Trends and Open Problems

Ankita Amburle¹ * and Ashok Kanthe²

^{1,2} Fr. Conceicao Rodrigues College of Engineering, Bandra-Mumbai, Maharashtra, India
ankita.amburle@fragnel.edu.in*
ashok.kanthe@gmail.com

Abstract. The present study is a bibliometric evaluation and an extensive analysis of the literature that has emerged concerning retrieval augmented generation as an educational tool in AI as published in Scopus from 2021 to 2025. The purpose of the analysis is to assess the researchers' potential in determining the flaws of today's educational system and improve learners' outcomes. We therefore analyzed more than 4500 research papers published in Scopus and identified more than 3400 keywords relevant to the topic of education. Furthermore, we attempted to find out how education is changing in relation to AI based on frequently used terms like LLM (1,338), learning (5,216), education AI (3,995), and RAG (324) and reached the conclusion that specific terms used in such applications as grading and tutoring do not receive much attention nowadays. The introduction of RAG-based Measurable Multimodal Educational Framework (MMEF) aimed at combining ethical principles, relation to teaching methods, and incorporated retrieval systems helped us clearly define the fields of study that have not been exhausted. The main part of the research incorporates three sections dealing with publishing trends, surveys, and keywords analysis. The research into retrieval augmented generation allows for educational interpretation of the ideas of LLM and refers to the basic methods of vector database technology and multimodal LLM system to demonstrate AI concepts used in educational domain. The research mainly focuses on revolutionary trends in "AI in education"; technology based learning represents the primary area of study.

Keywords: Retrieval-Augmented Generation (RAG), Artificial Intelligence in Education (AIED), Multimodal Learning Systems, Publication Trends

1 Introduction

Retrieval-Augmented Generation (RAG) is an AI system that generates text-based responses by combining retrieval from trustworthy data sources with language generation. It retrieves information from documents, databases, and knowledge bases, and then uses that information to generate answers instead of relying only on pre-existing model knowledge. As a result, RAG can reduce incorrect and ambiguous responses [1]. RAG can support educational institutions by providing tailored learning solutions and assessment tools that help students learn more effectively. It can also support medical professionals through clinical decision support and assist legal and

financial professionals with document review tasks [1], [2]. By grounding answers in current, trustworthy sources, RAG improves the reliability of AI outputs [4], [5]. In healthcare, it can help staff access precise medical information and clinical practice standards. In business, finance, and law, it can improve decision-making by providing access to relevant documents for customer support, risk assessment, and legal research. RAG can also support software development and research by enabling accurate access to code, documentation, and academic materials. Nevertheless, different applications of RAG have to be used since the processes of response generation and retrieval in different contexts differ from one another and are not always acknowledged in an obvious way [3]. While basic RAG systems are based on the use of semantic similarity to single out relevant documents to be sent to the generator, advanced RAG improves on the quality of retrieval by employing methods such as query rewriting, re-ranking, and hybrid retrieval using either dense or sparse techniques [4][5]. Modular RAG gives an opportunity for the user to choose necessary modules for retrieval, reasoning, and generation.[14], [15].

RAG is currently being implemented in educational institutions to create learning programs that offer flexible, personalized support, improved assessment techniques, and career planning services. RAG also assists teachers in curriculum development by enabling the creation of materials that align academic goals across subjects, provide a coherent academic flow, and address current industry needs using verified academic resources and up-to-date information on technological advances and professional competencies [16].

Retrieval-Augmented Generation (RAG) is an artificial intelligence approach that generates high-quality natural language by combining language generation with external knowledge retrieval. The solution solves the problems which exist in large language models because it produces false information and shows incorrect knowledge and operates without clear understanding in its pure generative model output. The RAG pipeline executes standard operations when its retriever uses user query information to find suitable content through semantic similarity assessment of selected knowledge sources which include vector databases and textbook collections and institutional repositories according to [10]. The LLM receives the retrieved context which enables it to produce answers based on current validated information and not just its trained knowledge. In educational settings, RAG provides personalized tutoring which creates curriculum content while enabling assessment through evidence-based methods that use established academic sources for their responses. Current research demonstrates that RAG-based systems improve factual accuracy and contextual relevance and explainability, which makes them suitable for educational settings that need both reliable results and correct teaching methods. The following points need a summary because they need emphasis. The discussion shows a significant move toward adopting AI-based educational systems, and Retrieval-Augmented Generation (RAG) has become one of the most important frameworks in recent studies. The solution solves the problems which exist in large language models because it produces false information and shows incorrect knowledge and operates without clear understanding in its pure generative model output. The RAG pipeline executes standard operations when its retriever uses user query information to find suitable content through semantic similarity assessment of selected knowledge sources which include vector databases and textbook collections

and institutional repositories according to [10]. The LLM receives the retrieved context which enables it to produce answers based on current validated information and not just its trained knowledge. In educational settings, RAG provides personalized tutoring which creates curriculum content while enabling assessment through evidence-based methods that use established academic sources for their responses. Current research demonstrates that RAG-based systems improve factual accuracy and contextual relevance and explainability, which makes them suitable for educational settings that need both reliable results and correct teaching methods. The following points need a summary because they need emphasis.

- Bibliometric evidence from Scopus-indexed literature indicates increasing integration of Large Language Models (LLMs) with external knowledge retrieval systems to improve accuracy and topicality in educational applications.
- The paper identifies key RAG components: retrievers, vector databases, embedding models, and multimodal LLMs as building blocks for knowledge-grounded AI education systems [12].

The following research questions guide this study:

- 1) What are the various publication trends over time in RAG research, what are the top journals, what are the most advanced countries and institutions in terms of this research, and how do different researchers collaborate in various countries?
- 2) What are the key themes in the literature on RAG and its keywords' co-occurrence, and how has the literature evolved over the years?
- 3) What are the less well-developed fields of RAG, based on keywords' low appearance frequency, lack of connectivity of the relevant research areas, and low numbers of publications?

2 Methodology

Recent developments in large language models (LLMs) and generative AI have significantly increased interest in multimodal Retrieval-Augmented Generation (RAG). This architecture enhances model outputs by retrieving and grounding information from multiple modalities (text, images, audio, and video). The publications covered by the dataset include examples of multimodal RAG development across several fields, including healthcare, robotics, education, human–AI interaction, multimodal perception, and information security.

Within the past five years there has been increasing research on AI and chatbot tutors in a variety of educational environments, particularly at the high school and university level. The latest review of AI based chatbot trainers for assisting students in an educational setting indicates most of the focus and attention from researchers is on undergraduate chatbots and general bulk use of AI, with only a limited amount of attention going towards primary school age children. While theorization has occurred through the contributions of authors such as Luckin et al. concerning scaffolding through AI, support should also be given toward the engineering of an AI chatbot for younger students, especially to ensure that their epistemological and psychological needs are adequately satisfied.

2.1 Data Collection and Screening

The Scopus search criteria focused on publications related to Retrieval-Augmented Generation (RAG) in education. The keyword “RAG (Retrieval-Augmented Generation)” was searched across all fields, yielding 4,610 publications published between 2021 and 2025. The search was restricted to research and review articles, excluded conference papers and book chapters, and was limited to English-language publications. The refined search which aimed to find RAG research conducted in higher education produced 1269 articles from the same time frame. The search term “multimodal RAG in education” resulted in 237 pertinent research documents.

The search was conducted using the following query formulation:

Topic Search = (“Retrieval Augmented Generation” OR “RAG”

OR “Multimodal RAG” OR “Multimodal Retrieval Augmented Generation” OR “Intelligent Knowledge Tutoring”) AND Publication year = (2021–2025) AND Document Type = (“Article” OR “Proceeding Paper”) AND Language = “English”

The methodological framework consists of three main components:

The first step of our research process requires us to conduct a literature search through Scopus databases and screen the results for relevant information. The second step of our research process requires us to create a topic modeling system together with a multimodal research assessment guideline system. The third step of our research process requires us to create educational applications which we will evaluate through synthesis and trend analysis.

This yielded 4,610 documents. Filters applied: – Document type: Articles and conference proceedings only (n=4,610 → refined to 4,500 relevant). Subject areas: Limited to Computer Science, Education, Engineering (manual screening excluded ~110 off-topic). Exclusion: Books, chapters, non-English (n<5%).

PRISMA flow: Initial 4,610 → screened titles/abstracts (n=4,500) → full-text relevance check (n=~1,500 for keyword extraction). Time range: 2021–2025 to capture RAG emergence post-LLM boom

2.2 Analysis Tools

This study adopts a systematic bibliometric and methodological review approach to examine the role of multimodal Retrieval-Augmented Generation (RAG) in the education sector [7]. Based on a bibliometric review, the following parameters were used in the review methodology Bibliometric visualization: VOSviewer v1.6.19 (min. keyword occurrence=10; co-occurrence network).

Keyword extraction: Scopus API + Python NLTK (3,400 terms; top: LLM=1,338, learning=5,216, education AI=3,995, RAG=324).

Thematic clustering: Manual + VOSviewer (5 clusters: NLP/QA, multimodal, ethics/learning outcomes)

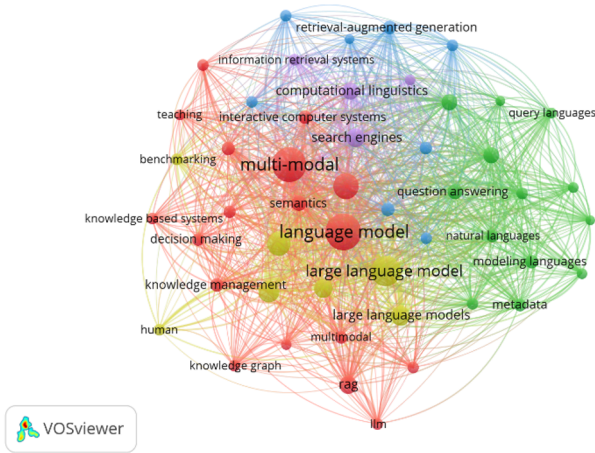


Fig. 1. Highlighting frequent words occurrences using VOSviewer with minimum keyword occurrence = 10

A total of 602 publications were obtained from the Scopus and Web of Science databases with the keyword Retrieval-Augmented Generation (RAG) through bibliometric techniques [11]. The application of visualization and cluster analysis through VOSviewer was used to identify some of the key themes including natural language processing, question answering systems, multimodal learning, and language models as shown in Fig 1. Five key research clusters were identified and especially focus on ethical considerations and learning outcomes of education. Out of them, 350 publications on ChatGPT were chosen, and 38 articles were out of the list of publications, which were reviewed in detail to determine new tendencies in the field of educational use. The analysis has also identified that the United States does the biggest share of research production and as such, the country has made significant contributions in RAG and NLP-based methods. The results underlined that requirement in RAG in education with multi modal inputs are a critical methodology to augment viability of space. Also, an analysis of 52 publications on multimodal RAG in education underwent a bibliometric analysis conducted by [6].

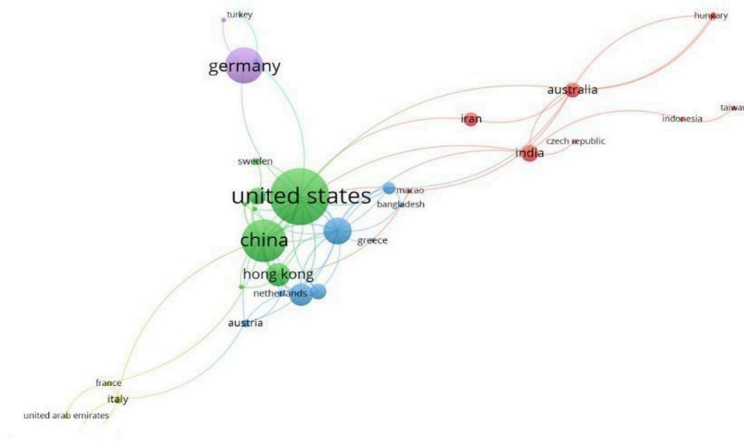


Fig. 2. Country wise Retrieval Augmented Generation RAG publication highlights

A bibliographical review of 560 articles of 2020-2025 about RAG Scopus was performed, and its effects on translation, coding and academic writing were identified. The analysis of the research was done using VOSviewer which tracks the frequency of keyword use, Co-citations and research networks. The bibliometric studies presented in Figure 1 and Figure 2 highlight the thematic and geographical coverage of the Retrieval-Augmented Generation (RAG) in the education study. Figure 1 depicts a keyword co-occurrence network that indicates that the main thematic groups composed of such terms as retrieval-augmented generation, large language models, natural language processing, question answering, and multimodal learning represent a significant focus on knowledge driven artificial intelligence and language-based learning applications. Conversely, education-specific functionalities, such as intelligent tutoring, grading, and assessment, are less densely connected, indicating that these research directions are not well researched now. In figure 2, the distribution by country shows that most publications are in the United States, then China and some European countries, thus representing a significant contribution of technologically developed areas. This distribution also shows the future but rather asymmetric international involvement in RAG based educational studies with little coverage of developing regions thus suggesting the presence of a lot of prospects of increased international cooperation and creation of regionally specific education applications. Fifteen hundred and eighty-eight publications on Retrieval Augmented Generation (RAG) in the Web of Science Core Collection (2020-2023) were analyzed using bibliometric analysis to identify the impact of this tool on translation, programming, and academic writing [12]. The research examined the co-occurrence of keywords, co-citation and research collaboration networks using VOSviewer. The findings have found that one of the key keywords with a high association with various thematic clusters is the language model. The United States, China, and Australia dominated the country level collaboration, with the institutional analysis showing China on general leadership in this collaboration with three out of five positions at the

top in overall link strength. The results present some useful information on the possibilities of RAG implementation in the sphere of education. In another related research, a mixed-method bibliometric analysis and systematic literature review was conducted to analyze 120 publications assessing the application of OpenAI ChatGPT in education [10]. The study found significant contributors, dominant themes of research, and new directions of research, where there was a swift increase in publications between 2021 and 2024. The network and keyword analyses identified several underdeveloped areas, among which are the challenges, pedagogical practices, and knowledge building. In addition, an example of a bibliometric analysis of RAG as a future research area in education 51 documents indexed in the Web of Science Core Collection in 2023 [14]. Through bibliometric measures and network analysis, the analysis of citation patterns, collaboration structure and temporal development of the research topic were done and offered a broad picture of development of the area. Additionally, the authors of the article clearly investigated the usefulness of multimodal RAG (38) research publication through Scopus, in education process identifies various inputs to improve the learning capabilities of students. Out of preliminary searches of 520 journal publications anchoring on RAG and multimodal RAG, 120 were chosen to study. The appearance of a bibliographical keyword in publication that Chatbot is mentioning in one of the publications is an education-related content.

3. Proposed Framework

A. RAG-Integrated Multimodal Educational Framework (RIMEF)

The RIMEF Model is the primary contribution of this project. The RIMEF Model was developed for the purpose of providing a more pedagogically sound approach to conventional LLMs. While traditional generative models rely only on parametric dimensions of knowledge, RIMEF enforces three layers of interactivity through a pipeline:

1. Ingestion Layer: The first layer is the Ingestion Layer where all the institutional knowledge, such as curricula standards, textbooks, lecture notes are curated into a Vector Database using some type of dense embedding technique (e.g., Ada or Sentence-BERT).
2. Retrieval & Ranking Layer: The second layer is the Retrieval & Ranking Layer, which is an iterative module that takes the Hybrid Retrieval (composed of sparse keyword searching and dense semantic vector searching) to obtain the most relevant educational resources from the institution.
3. Synthesis Layer: The final layer is the Synthesis Layer, which is where retrieved chunks of knowledge are used as contextual inputs to an LLM for generation with constraints from a system prompt to ensure pedagogical accuracy and explainability of the generated content.

Figure 3 depicts a retrieval-augmented generation (RAG) system for an education chatbot. The student submits a learning request to the AI tutor, which is processed through the Retrieval phase, during which a user's request is transformed into a learning object (the subject of learning), symptoms of interest (the specific skill(s) the student is acting on) and verified data from which to create the response. The Retrieval phase uses the user's request to identify contextual and/or goal-based sources or databases that contain previously created responses to the same learning

request. Note: The RAG process has no data/knowledge of where the user is located or what is available to them as a knowledge source. Also, RAG uses knowledge sources outside of the traditional confines of an institution of higher education that contain verified data (i.e., databases that contain confirmed/certified knowledge content). Once identified, the verified source/learning object(s) will be automatically sent to the Generation phase where the verified content will then be integrated with the knowledge that exists within the user's current query to produce an accurate and contextually relevant answer. The system is capable of interpreting the information received through RAG to generate multiple outputs (multimodal). Therefore, the generation of explanations will yield output formats such as text-based explanations (i.e., word-based), graphic-like (i.e., diagram-based) and video-based (where video or audio are also used) explanations). Finally, the output produced will be provided in a format that allows for the integration of data to create an education response based upon the learner's previous learning experience; thus, enabling improved conceptual comprehension, adaptive learning, and learner-engagement [8]. Overall, these illustrations illustrate how RAG-based education chatbots provide verification-based knowledge and contextualize the learning experience through verified sources of knowledge while providing learners with personalized, interactive learning experiences.

4. Results

Large Language Models have several limitations, including fixed training data, and they can produce unverifiable or outdated content. Retrieval-Augmented Generation (RAG) mitigates these problems by grounding responses in externally stored, domain-specific knowledge. This integration improves factual accuracy, transparency, and reliability of generated outputs. Consequently, RAG is particularly useful when evidence-based, up-to-date, and explainable intelligence is required in research applications.

Table 1 presents a comparative outline of the most recent studies in the field of Retrieval-Augmented Generation (RAG) techniques applied for various areas of study. It contains reference to important points of research such as year of publishing, usage of RAG technologies, utilized datasets or benchmarks, and areas of technology implementation. Some papers are devoted to educational support systems, while others show a broader perspective on the application of RAG techniques in healthcare, industrial automation as well as business analytics. This comparison provides an overview of the modifications of RAG architectures to special domain needs while paying special attention to improving factual accuracy, contextual relevance, and reliability of AI-based systems.

Table 1. Summary of RAG-based applications across domains

Sr. No.	Paper (Year)	Use Case with RAG	Datasets / Benchmarks Used	Application Area
1	Pabon-Mendoza et al. (2025)	Academic assistant retrieves student information, notes, and FAQs	University documents, student queries	Educational support systems
2	Xie et al. (2024)	Knowledge retrieval for Visual Question Answering (VQA)	VQA-v2, OK-VQA, GQA	Vision–language reasoning, diagrams
3	Akbar et al. (2025)	Conversational memory retrieval for chat systems	MultiWOZ, conversational logs, agriculture documents	Conversational AI, agriculture
4	Patel & Sharma (2025)	Secure retrieval of healthcare records	MIMIC-III, synthetic EHR datasets	Healthcare, medical AI
5	Woesle et al. (2025)	Use of RAG to reduce hallucinations	TruthfulQA, NQ, HotpotQA, Wikipedia	AI safety, hallucination detection
6	Li et al. (2025) (SAMACR3-MED)	Multimodal RAG for engineering PDFs (images + tables)	Engineering manuals, technical PDFs	Engineering document AI
7	Wen et al. (2025)	RAG for industrial knowledge retrieval	Industry workflows, SOPs, enterprise documents	Knowledge management, enterprise AI
8	Chen et al. (2025)	RAG for interactive industrial knowledge systems	Maintenance logs, machinery datasets	Industrial automation
9	Praneeth et al. (2025)	RAG for customer feedback summarization	Amazon/Flipkart reviews	Sentiment analysis, business analytics

10	Saha et al. (2025) (QuIM-RAG)	RAG-based QA with inverted question matching	SQuAD, NQ, MS-MARCO, HotpotQA	QA systems, educational bots
----	-------------------------------	--	-------------------------------	------------------------------

Suitable publications were obtained from Scopus by means of appropriately selected search strings employed in the combination of multi-modal learning, RAG, and educational contexts. Then, based on inclusion and exclusion criteria, we processed the studies by analyzing publications tendencies, main research areas, methods used, and their practical application. Besides, we analyzed a suggested multimodal RAG solution that applies text, images, and structured data retrieval in combination with large language models to solve problems in education. Finally, we carried out descriptive statistics and thematic analysis of the data.

A student’s question initiates RAG, activating the semantic retriever that extracts pertinent pieces of content from the vector database. A re-ranker follows next to eliminate irrelevant noise, using cosine similarity as a scoring technique to select the most significant content. Next, the content is input into the LLM together with a strict prompting template like “Respond but only on the basis of this context; tell if it is insufficient.”

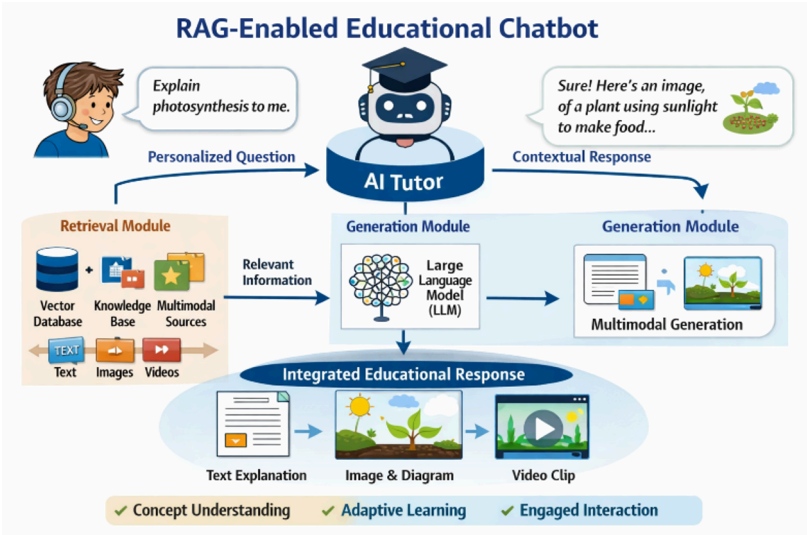


Fig. 3. RAG-Integrated Multimodal Educational Pipeline: Student query → Hybrid Retrieval (text/images) → LLM Generation → Multimodal Response (text/diagram/video) with teacher governance loop.

5. Discussion

The broad overview of RAG, as well as RAG + generative, multimodal and LLM solutions, are illustrated with examples from various industries including healthcare, autonomous driving, mapping, data-driven decision-making, and education. Based on this research, a RAG-based educational chatbot architecture demonstrates successful integration of retrieval and generative models through personalization/multimodal

learning. This implementation uses vector databases and curated education knowledge bases to ensure accuracy and currency of the generated response. Multimodal contribution (text, images, and videos) is an answer to the needs of various types of learners and promotes conceptual knowledge, which is especially valuable in the case of complex or abstract topics. This architectural view supports the view that RAG is not an augmentation of LLMs, but a paradigm of design of future educational systems. The other significant finding is that the RAG approaches can be translated in other fields, especially in an educational setting out of healthcare and finance. The methods which were initially created to minimize hallucination, compliance, and decision support are now being modified to serve educational purposes. This cross-domain effect serves as one of the indicators of the maturity of the RAG as a generalizable framework and the necessity of education-specific measures of evaluation, ethical aspects, and long-term deployment research. The study presents an extensive review of the recent research on Retrieval-Augmented Generation (RAG) in educational artificial intelligence through synthesizing bibliometric findings, architectural commentary, and cross-domain comparisons. The results show that the development of AI-based educational studies is expanding rapidly, and RAG became one of the key models to overcome the shortcomings of classical Large Language Models (LLMs), especially regarding factual underpinning, situational applicability, and interpretability. According to the bibliometric analysis of publications Scopus indexes, it is possible to note the rise in the popularity of such terms as learning, education AI, LLMs, and RAG, which proves the shift in the paradigm toward knowledge-based and adaptive learning systems.

A. Challenges Although Retrieval-Augmented Generation (RAG) is increasingly being used in educational systems, there are various challenges that inhibit its effective and large-scale application. The reliance on quality and edited sources of knowledge is one of the main issues. The quality, relevance, and completeness of educational content used in the implementation of RAG based educational tools are very sensitive when it comes to the performance of these tools. The other major difficulty is the retrieval accuracy and relevancy congruence. Even though the contextual matching in a vector-based semantic search is enhanced, it is possible that irrelevant or conceptually close content is retrieved which is not exactly matching the intent of the learner or the purpose of the curriculum. The problem is more acute in the educational setting, and conceptual accuracy and pedagogical correctness is crucially needed. RAG systems also have complex computations and resource demands. Multimodal RAG systems that combine text, pictures, and video materials demand intensive computing, memory, and storage space, which cannot be easily deployed in resource-limited institutions of learning. Real time response generation also adds to latency and overhead of systems. The issues of privacy and ethics are especially severe in education. RAG-based systems are frequently based on student data, academic history, and interaction history to offer individual learning. The obligation to adhere to data protection laws and to ensure that the data is not accessed by unauthorized parties are the tasks which must be continued. Explainability and trust among teachers may be a barrier to adoption. Although RAG can enhance the factual grounding, the process of retrieval and generation is not always clear to the educators. Trust could be established only based on decipherable retrieval systems and explicable AI products which should meet the requirements of the curriculum and instructional goals. The issues of privacy, ethics

and data governance are also significant, especially when retrieval-augmented generation (RAG) systems work with student data to personalize them. Although the literature provides evidence of the existence of ethical concerns, the findings demonstrate that there is scanty empirical research on the governance mechanisms. Using or accessing sensitive information of learners without permission may compromise trust or adherence to regulation by the institution. Such risks may be alleviated by using privacy-protecting retrieval techniques, anonymizing student data, accessing this data through secure means, and adhering to educational standards of data protection. Hallucination is a continuing issue in educational systems that use large-scale language models, which can give fluent but false or unprovable responses. The reason why this phenomenon has been largely due to the over dependence on parametric knowledge and poor grounding in the generation, which is a major challenge that creates enormous issues in learning situations that demand exactness and conceptual legitimacy. Hallucinations have consequences, which include the spreading of misinformation, loss of trust among the learners and potential lack of congruency with the curricular objectives.

5. CONCLUSION

The paper is a systematic review of Retrieval-Augmented Generation (RAG) in educational artificial intelligence, which examined 10,020 publications by Scopus that are four to five years old. The results show a solid move towards the systems of learning with higher levels of the implementation of LLM and the retrieval to enhance the facts, context, and their personalization. The discussion reveals that although educational chatbots and learning support systems are well researched, other systems like intelligent tutoring, grading, and assessment have not been developed well. The suggested RAG-based educational chatbot architecture illustrates the successful approach to the multimodal integration of the material to support adaptive and student-centered learning processes. Altogether, the research makes RAG a transformative facilitator of AI in education and sets definite paths of the future research. Further studies are needed which should not merely be bibliometric exploration but rather empirically prove the use of RAG-based educational systems. The possible directions are to apply curriculum-based RAG architecture, measure retrieval and generation performance with reference to learning levels and assess the effects of RAG-enabled tools on learning, assessment accuracy and educator confidence. Furthermore, analysis of multimodal and non-English educational data can help to gain a more in-depth or wholesome insight of how RAG applies to various learning environments.

References

1. J. Pabon-Mendoza, A. López, and M. Rivera, "Design of an academic assistant using retrieval-augmented generation for student support," *IEEE Access*, vol. 13, pp. 1–12, 2025.
2. Y. Xie, Z. Liu, and H. Wang, "Retrieval-augmented visual question answering for multimodal reasoning," *IEEE Transactions on Multimedia*, vol. 26, pp. 1453–1465, 2024.
3. M. Akbar, S. Rahman, and A. Khan, "Conversational memory retrieval using retrieval-augmented generation for domain-specific chat systems," *Expert Systems with Applications*, vol. 238, pp. 121987, 2025.

4. R. Patel and S. Sharma, "Secure retrieval-augmented generation for healthcare record access," *Journal of Biomedical Informatics*, vol. 152, pp. 104112, 2025.
5. T. Woesle, K. Muller, and A. Singh, "Reducing hallucinations in large language models using retrieval-augmented generation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 4, pp. 4821–4829, 2025.
6. Y. Li, P. Zhang, and L. Chen, "SAMAC-R3-MED: Multimodal retrieval augmented generation for engineering and medical documents," *IEEE Transactions on Artificial Intelligence*, vol. 6, no. 1, pp. 77–89, 2025.
7. J. Wen, R. Gupta, and M. Zhou, "Retrieval-augmented generation for industrial knowledge management systems," *Knowledge-Based Systems*, vol. 295, pp. 111842, 2025.
8. H. Chen, D. Park, and S. Lee, "Interactive industrial knowledge systems using retrieval-augmented generation," *IEEE Transactions on Industrial Informatics*, vol. 21, no. 3, pp. 2456–2467, 2025.
9. K. Praneeth, A. Mehta, and V. Rao, "Customer feedback summarization using retrieval-augmented generation," *Information Processing Management*, vol. 62, no. 2, pp. 103567, 2025.
10. S. Saha, R. Banerjee, and P. Mitra, "QuIM-RAG: Inverted question matching for retrieval-augmented question answering," *ACM Transactions on Information Systems*, vol. 43, no. 1, pp. 1–28, 2025.
11. S.-V. Oprea and A. Bara, "Transforming education with large language models: Trends, themes, and untapped potential," *IEEE Access*, vol. 13, pp. 87292–87310, 2025, doi: 10.1109/ACCESS.2025.3570649.
12. F. Liu, "AI-empowered pedagogical reform in power electronics: A practical approach based on retrieval-augmented generation and virtual simulation," in *Proceedings of the 2025 2nd International Symposium on Artificial Intelligence for Education (ISAIE 2025)*, Changsha, China, Sept. 19–21, 2025, pp. 149–153, doi: 10.1145/3775073.3775099.
13. K. N. Khanghah, Z. Chen, L. Romeo, Q. Yang, R. Malhotra, F. Imani, and H. Xu, "Zero-shot anomaly detection in laser powder bed fusion using multimodal retrieval-augmented generation and large language models," *Journal of Mechanical Design*, vol. 148, no. 7, Art. no. 072001, 2026, doi: 10.1115/1.4070585.
14. S. Pozdniakov, J. Brazil, O. Poquet, S. Krusche, S. Berrezueta-Guzman, S. W. Sadiq, and H. Khosravi, "From knowledge gaps to learning opportunities: Leveraging student questions and dual use of generative AI to support student learning at scale," *Computers and Education: Artificial Intelligence*, vol. 10, Art. no. 100509, 2026, doi: 10.1016/j.caeai.2025.100509.
15. A. Vidyarthi, K. M. Singh, and D. S. Moirangthem, "SageRAG: Query rewriting for retrieval enhancement and retrieval-augmented generation for grounded responses in AI research assistance," *Expert Systems with Applications*, vol. 308, Art. no. 131160, 2026, doi: 10.1016/j.eswa.2026.131160.
16. A. Amburle, V. Shelake, A. Kanthe, and A. Malkar, "Aspect Based Call Analysis of Marathi Regional Language in India Using Machine Learning

- Approach,” in *Proc. 2024 Int. Conf. on Innovation and Novelty in Engineering and Technology (INNOVA)*, 2024, doi: 10.1109/INNOVA63080.2024.10847022.
17. N. Gertler and R. Sabnekar, “Developing the Luciano Floridi (LuFlot) bot: An accessible AI chatbot trained on a philosopher’s manuscripts,” *Minds and Machines*, vol. 36, no. 1, p. 7, 2026, doi: 10.1007/s11023-026-09761-4.
 18. K. Sharma, A. Amburle, V. Shelake, and A. A. Chauhan, "Development of an AI Based Priority Formation Mechanism in Criminal Law," in *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Pudukkottai, India, 2024, pp. 1542–1548, doi: 10.1109/ICACRS62842.2024.10841680.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

