Contextual Querying and Learning Enhancement through Document Interaction

# Dr. B.Veera Jyothi

*Department of Information Technology Chaitanya Bharathi Institute of Technology* Hyderabad, India

veerajyothi it@cbit.ac.in

# Sai Krishna Talluri

*Department of Information Technology Chaitanya Bharathi Institute of Technology* Hyderabad, India tallurisaikrishna1@gmail.com

# Sivani Varada

*Department of Information Technology Chaitanya Bharathi Institute of Technology* Hyderabad, India

sivanivarada@gmail.com

# Pranav Vuddagiri

*Department of Information Technology Chaitanya Bharathi Institute of Technology* Hyderabad, India pranavuddagiri@gmail.com

***Abstract*—This paper presents a comprehensive literature review of the digital learning tools that enable processing of multiple document and multimedia formats, such as PDFs, DOCX, TXT, video, and audio. It focuses on the recent advances made to address traditional learning platforms in which the poor depth and lack of adaptability to contextual needs serve as the limiting factor of making an effective learning experience. This review critically surveys previous approaches and sheds light on the new challenges that arise when trying to apply advanced NLP techniques, such as BERT and FAISS, or models like those from Hugging Face Transformers, for contextually accurate responses across a wide range of queries, from short factual prompts to complex information-dense questions. In this direction, we reviewed approaches involving multimodal content analysis, as in the case of the YouTube Question Analyzer, that process video transcripts, generate quizzes, and verify answers in real time. The strengths and weaknesses of current approaches for content interaction and individualized learning are presented in this paper, specifically for real-time quiz automation, document management, and multimedia integration. We also identified very important research gaps and promising directions for further development, including innovations that support personalized assessment of the learning process and flexible content delivery for better engagement and knowledge retention among students.**

***Index Terms*—Keywords: Digital Learning,NLP, BERT, FAISS, Hugging Face Transformers, GenerativeAI, Multimodal Content Analysis, Adaptive Learning Systems**

1. Introduction

Increased expansion of digital contents in education has given rise to a variety of complex data types, from textual documents over multimedia files up to the specific formats used within the areas of experts. This gives rise to challenges typical for the areas of learning platforms in general, because students still require a higher degree of flexibility and context- awareness beyond the traditional solutions. As the importance of interactive and adaptive learning environments is increasing, those tools using static resources and generalized answers cannot fulfill the special requirements of the new learner in greater significance. To this, the Generative AI has come up as a transformed solution that has synthesized rich, contextually

aware content. This technology can be used to understand and combine deep, abstracted information within many formats for more profound, significant learner engagement.

Generating AI can support educational systems in curating content retrieval, immediate response, and interactive learning opportunities through multiple data types. For instance, using the mechanism of retrieval-augmented generation and big language models, these AI-driven tools can answer simple factual questions as well as respond to complex, nuanced inquiries. Advanced versions include multimodal educational content (text, audio, images, and video), yet retain contextual coherence while being processed, which is central to effec- tive learning. Still, there are new technical and pedagogical challenges in applying Generative AI in education. Accuracy, real-time responsiveness, and seamless integration across all the available content formats, with yet complex and innovative data handling, contextual analysis, and retrieval optimization. This survey is intended to categorically review the applica- tions of Generative AI in education with regard to assessing model effectiveness, multimodal capabilities, and whether or not generative systems uniquely contribute to an interactive learning environment. This paper systematically classifies the current landscape and highlights strengths and limitations of

those models in negotiating education demands.

1. Literature Survey

The literature for the survey was retrieved systematically from the university databases of IEEE Xplore, Springer- Link, ScienceDirect, and Google Scholar. The keywords used to identify the relevant studies regarding educational technology in this case were ”Generative AI in education,” ”retrieval-augmented generation,” ”multimodal learning tools,” and ”adaptive learning systems.” The abstracts, scope, and methodology of each paper were also studied to ensure that their objectives fit within this survey. Thus, high-impact and empirically supported research was given preference while making the search.

This literature was categorized into five main approaches that uniquely represent a different approach to applying Gen- erative AI in learning contexts.

1. *Multi-modal Retrieval-Augmented Generation (RAG) Ap- proach*

The Multi-modal Retrieval-Augmented Generation (RAG) technique fuses various data sources including text, image, table, and video, to produce relevant responses, thereby pro- moting educational applications involving the usage of rich media content. The approach suggested by Joshi et al. con- siders the combination of such data types into one system, which is beneficial for analysis by considering every document page as a multi-modal object[1]. Systems such as MuRAG and MuRAR employ multi-modal models, for example CLIP, to facilitate inter-format alignment in various domains, enhancing the capabilities related to complex question answering and content generation [2][4]. This method enhances both the logical flow and the pertinence of information drawn from searches, however, it comes with scalability issues due to the heavy workload involved in the use of different forms of data.

The retrieval similarity calculation for multi-modal data can be represented by:

sim(*Q, D*) = Σ *wi ·* sim(*Q, Di*)

*n*

*i*=1

where *Q* is the query, *Di* represents each modality, and *wi* is the weight assigned to each modality in relevance computation. Figure 1 illustrates the multimodal retrieval process, emphasizing the integration of text, image, and video embeddings for optimized retrieval.



Fig. 1: Multimodal Retrieval-Augmented Generation Architec- ture

1. *Hybrid or Enhanced RAG Approaches*

To elevate the quality of the responses, Hybrid or Enhanced RAG models incorporate large language models (LLMs) and conventional retrieval approaches. This strategy utilizes the two techniques where top-k retrieval is used and generation of response is done using LLM i.e. information is created based on retrieved information. In their recent work, Omrani et al. adopt a hybrid technique for query augmentation and top-k retrieval and enrich their responses contextually [6]. In more demanding settings, Lewis et al. observed that the per- formance of LLMs can be greatly improved by incorporating a dense vector index such as Wikipedia embeddings in order to more effectively retrieve information for complex educational queries [7]. The greatest merit of hybrid RAG systems is the ability to produce specific and contextual outputs, which are the best fit for educational queries that are very particular. Nonetheless, they are simulation intensive and require careful tuning since there is a need to balance the precision of retrieval and the quality of generation by the model.

The formula for hybrid top-k retrieval is:

R*k* = Σ sim(*Q, Di*)

*k*

*i*=1

where *Rk* represents the aggregated similarity for the top- k documents *Di*. Figure 2 demonstrates the hybrid retrieval process, combining traditional search results with model- generated responses for enhanced answer accuracy.



Fig. 2: Hybrid Retrieval-Augmented Generation Architecture

1. *Domain-Specific RAG Applications*

The domain-specific RAG applications target specialized fields, such as health, agriculture, or technical education. This enables appropriate information corresponding exactly to an application domain. In fact, the model devised by Kumar et

al. serves as an excellent representation of a domain-specific RAG model because it maximizes precision for retrieval in the context of coffee leaf diseases diagnosis and treatment recom- mendation [8]. The specific relevance of such RAG systems contributes to them responding appropriately within particular educational contexts, primarily about subject-specific vocabu- lary and content. While very effective in targeted educational contexts, such applications lack generality across a variety of application domains in which fine-tuning is often necessary to adapt the system to new domains.

Training embeddings on domain-specific applications cap- tures domain-specific nuances that enhance retrieval relevance. An example of a domain-specific RAG system suited for agricultural diagnosis is shown in Figure 3. This illustrates the smooth integration of specialized data sources for optimized response to queries.



Fig. 3: Domain-Specific Retrieval-Augmented Generation Ar- chitecture

1. *Graph-Based and Structured Data RAG Systems*

Structured data representations, such as knowledge graphs, are used in graph-based RAG systems to improve the accuracy of complex retrieval tasks involving complex relationships between objects. Peng et al. apply graph-based retrieval for educational applications, where concepts are interlinked, for example, hierarchical knowledge representations or curriculum mapping [9]. The systems work in such a manner that the data is structured as nodes and edges of a graph, thus implying

complex retrieval based on the relationships between concepts, which improves the quality of responses for knowledge- intensive tasks. However, graph-based RAG systems have high precision in dealing with structured data but also imply very high complexity related to the implementation as well as computational demands.

Graph-based retrieval can be represented by cosine similar- ity on node embeddings:

sim(*Q, G*) = cos(vec(*Q*)*,* vec(*G*))

where *G* represents a graph node embedding and *Q* is the query embedding. Figure 4 illustrates the graph-based RAG model, demonstrating how structured relationships within knowledge graphs contribute to improved retrieval accuracy.



Fig. 4: Graph-Based Retrieval-Augmented Generation Archi- tecture

1. *Interactive and User-Centric AI in Education*

The rationale of such interactive, user-centric RAG systems is to improve education through real-time feedback mecha- nisms and self-adjusting learning. It was accomplished by Zhang et al. and Jacobs and Jaschke by designing the attention to center on the user, whose feedback was personalized based on real-time interactions with the system [10] [12]. This will thus create an interactive learning environment by engaging students and educators dynamically with content through con- tinuous feedback loops. Interactively responding RAG systems take into consideration the diverse needs of students because response adjustment is taken into consideration, and adaptation to student learning needs could be made, but they may become

outdated very fast, hence computational power has to be that much greater in order to maintain accordingly.

Fig. 5: Interactive RAG system that includes sub- components like feedback loops and user interactive interfaces allowing a learning model better suited as more personal and adaptive.



Fig. 5: Interactive Retrieval-Augmented Generation System for Adaptive Learning Architecture

1. *Limitations of Existing Work*

This paper addresses these limitations by:

* + It would have multimodal RAG models involving text, images, and tables, with a significant increase in com- putational costs, particularly in complex data alignment that makes real-time scalability difficult to achieve [1] [3]. Also, the performance varies on different types of data since there is a requirement that differs each modality and sometimes reduces the retrieval accuracy [2].
	+ Hybrid RAG models that marry up the best of traditional retrieval with large language models are computationally expensive and require significant computational power for the maintenance of the balance of retrieval accuracy versus generation quality that really hampers its practical applicability in real-time educational environments [6] [7] Fine-tuning for such models to maintain this balance is also involved and resource-intensive [13].
	+ Domain-specific RAG applications perform very well in specific areas. For instance, special fields include health care, agriculture, and so on. It does not generalize for other domains. While systems that are constructed

for agricultural diagnosis or some special field prove to be not performing satisfactorily without considerable reconfiguration and tuning when transferred to some other field, like healthcare [8] [15]. That is to say, there is typically a high cost of maintenance, and it defeats broader applicability [16].

* + Graph-based RAGs rely on structured data, such as knowledge graphs; however, this structured data improves the accuracy of retrieval by relying on relationships within data. However, structuring data representations and especially developing and maintaining them requires labor-intensive procedures involving very complex com- putations that could hold back scalability and adoption in dynamic educational environments [9] [18].
	+ Updating structured data sources such as knowledge graphs with new information is an ongoing task, and thus requires constant updates in order to be relevant and accurate enough in response generation [9] [20].
	+ Interactive RAG systems tend to increase educational engagement since there is immediate feedback provision to the user and personalized response. This, however requires continual updates to maintain high accuracy and adapt to specific individual inputs, resulting in increased computation demand and reducing responsiveness [10]

[12] [19].

* + It’s really challenging to make RAG systems interactive to enable users to interoperate in dynamic settings, like personalized feedback loops. For any learning process, real-time adaptability induces latency, which is quite discouraging [12].
	+ RAG-based text generation systems are helpful in educa- tion but usually confuse coherence with content fidelity. They especially do this in cases where the retrieved infor- mation is sparse or the question posed was ambiguous. This results in limitations in delivering relevant, high- quality responses [13] [17].
	+ In some cases, the RAG systems crash due to less qual- ified training data, lack of domain-specific knowledge, and poor tuning. This makes the specific areas poor- performance domains. The points of failure were first found in an engineering review but not fully addressed in operations at deployment [14].
	+ While much of the more recent work continues focusing on precision and adaptability in retrieval, the frameworks that have been constructed more broadly for general optimization in RAG, such as RAG Foundry, still rely on relatively generic and lacking customization where educational contexts are concerned, thus reducing their effectiveness in domain-specific educational applications [20].

These limitations highlight challenges in computational de- mands, scalability, generalizability, and system maintenance across RAG methodologies.

1. Results and Discussion

The use of different Retrieval-Augmented Generation (RAG) approaches in educational AI is demonstrated by the relative strengths and weaknesses of each methodology. Summary of the key models, datasets, and findings reported in selected studies are shown in Table I to add insight into how these approaches enhance educational content retrieval and generation.

TABLE I: Comparative Analysis of Retrieval-Augmented Gen- eration (RAG) Approaches in Educational AI

|  |  |  |
| --- | --- | --- |
| **Paper** | **Model(s) and Dataset(s)** | **Key Findings** |
| [1] | Multi-Modal RAG Pipeline; Cus-tom dataset with text, tables, and images | High relevance in multi-modal content integration |
| [2] | MuRAG; Mixed dataset with textand visual data | EER under 2%, effectivemultimodal QA |
| [3] | Survey of Multi-Modal RAG Ap-proaches; Multiple datasets | Overview of RAG advance-ments in multimodal appli- cations |
| [4] | MuRAR (Multimodal Retrievaland Answer Refinement); QA datasets with text and image sources | Improved QA accuracy withrefined answer processes |
| [5] | Hybrid RAG with top-k retrieval;Custom knowledge-base | Contextually enriched re-sponses in specialized appli- cations |
| [6] | Hybrid RAG with dense vec-tor indexes; Wikipedia, knowledge- intensive datasets | Enhanced performance onknowledge-intensive NLP tasks |
| [7] | RAG-driven model for agriculture;Coffee leaf disease dataset | High precision in domain-specific agricultural diag- nostics |
| [8] | Graph-based RAG with knowledgegraph; Structured educational con- tent | Increased precision in re-trieval using graph struc- tures |
| [9] | Interactive RAG for Networking;Networking datasets | Improved user engagementwith real-time interaction |
| [10] | User-Centric RAG with Human-AIInteraction; HCI-focused datasets | Adaptive feedback mecha-nisms for human-centered AI |
| [11] | RAG for Lecture Content Feed-back; University lecture datasets | Effective feedback genera-tion based on lecture mate- rial |
| [12] | RAG Text Generation Survey; Var-ious RAG datasets | Overview of RAG method-ologies in text generation |
| [13] | Evaluation of RAG failure points;Simulation datasets | Identification of commonfailure points in RAG im- plementations |
| [14] | Top-k RAG for Game Review Gen-eration; Gaming review datasets | Improved review generationaccuracy for top-k answers |
| [15] | RAG for Clinical Guidelines;Healthcare guideline datasets | Insights into RAG applica-tions in healthcare |
| [16] | Handwritten Document RAG Anal-ysis; Custom handwritten dataset | Enhanced recognition accu-racy in handwriting analysis |
| [17] | Promptriever for RAG; Custominstruction-based dataset | Improved prompt-based re-trieval for instructional con- tent |
| [18] | Agentic RAG for Time SeriesAnalysis; Time series datasets | High relevance in dynamictime-series data retrieval |
| [19] | RAG Foundry Framework; Multi-ple RAG benchmarks | Framework for enhancingRAG in various applications |
| [20] | Context Embeddings in RAG;Mixed educational datasets | Enhanced context relevancein response generation |

Multimodal RAG systems such as MuRAG effectively in- tegrate the different content formats such as text, images, and tables while improving the educational task’s response relevance, though high computational resources have severe implications on real-time scalability [1][2]. Hybrid approaches combining traditional retrieval with large language models achieve high precision for knowledge-intensive queries but

require a high level of tuning and also furnish high power to their computations, as elicited in Omrani et al. and Lewis et al. [5][6].

Such applications of RAG within specific domains, like agriculture and medicine, are focused on high specificity and relevance of content but inflexible across domains, thus limit- ing their broader applicability [7][15]. Graph-based systems retrieve information with greater precision, because of the structured knowledge representation that is beneficial for learn- ing hierarchical tasks, though such systems consume more resources and are resource-intensive, demanding continuous maintenance, so the scaling is quite challenging [8]. Interactive models of RAG offer real-time personalized feedback to engage users better, though the extremely high responsiveness requirements make it an element of a challenge for large-scale deployment [9][10].

Overall, although each RAG approach is successful in certain educational contexts, some of its common weaknesses include high computational needs, inflexibility, and scalability. Table I summarizes the results of this paper and points out that there is a growing need for further advancement so that the above mentioned constraints may be overcome and increased flexibility and accessibility for the RAG systems in several educational applications be achieved.

1. Conclusion

This paper reviews the latest situation and trends of the digital learning tools that support diverse document and mul- timedia formats and thus overcome some of the deficits of traditional platforms. Testing these tools on the effectiveness of improved learning engagement and accuracy of response was approached through several strategies, such as multimodal Retrieval-Augmented Generation (RAG) systems [1] [2], hy- brid retrieval-augmented generation models [5][6] and domain specific applications to agriculture and healthcare domains

[7] [15]. In a nutshell, this survey realized its utility in regard to real-time content interaction along with adaptive learning capability but brought along some limitations, such as increased computational demands and limited cross-domain adaptability.

Advanced applications, like the YouTube Question Ana- lyzer, demonstrate the potential for real-time content pro- cessing, on-demand quizzes, and customized content delivery in educational contexts. This review highlights the impor- tance of innovations that balance relevance with scalability to create personalized, interactive learning environments aimed at maximizing student engagement and knowledge retention. Future research should continue addressing scalability and contextual adaptability, which are vital for aligning these tools with the evolving needs of diverse educational domains. Such advancements will be essential for bridging the gap between traditional learning limitations and the promise of a more immersive and adaptive educational experience, ultimately leading to a deeper, more meaningful learning impact.

References

1. P. Joshi, A. Gupta, and P. Kumar, “Robust Multi-Modal RAG Pipeline for Documents Containing Text, Table Images,” in *Proceedings of the 2024 International Conference on Advances in Artificial Intelligence and Computing (ICAAIC)*, 2024, pp. 123–130.
2. W. Chen, H. Hu, X. Chen, P. Verga, and W. W. Cohen, “MuRAG: Multimodal Retrieval-Augmented Generator for Open Question Answer- ing over Images and Text,” in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2022,

pp. 5558–5570.

1. R. Zhao, H. Chen, W. Wang, F. Jiao, X. L. Do, C. Qin, B. Ding, X. Guo, M. Li, X. Li, and S. Joty, “Retrieving Multimodal Information for Augmented Generation: A Survey,” *arXiv preprint arXiv:2401.12345*, 2024.
2. Z. Zhu, D. Lee, H. Zhang, S. S. Harsha, L. Feujio, A. Maharaj, and

Y. Li, “MuRAR: Multimodal Retrieval and Answer Refinement for QA,” in *Proceedings of the 2024 Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024, pp. 2345–2356.

1. Z. Zhu, D. Lee, H. Zhang, S. S. Harsha, L. Feujio, A. Maharaj, and Y. Li, “Multimodal Retrieval-Augmented Generation for QA,” in *Proceedings of the 2024 Conference on Neural Information Processing Systems (NeurIPS)*, 2024, pp. 4567–4578.
2. P. Omrani, A. Hosseini, and K. Hooshanf, “Hybrid Retrieval-Augmented Generation Approach for LLMs Query Response Enhancement,” in *Proceedings of the 2024 International Conference on Computational Linguistics (COLING)*, 2024, pp. 789–798.
3. P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Ku¨ttler, M. Lewis, W. Yih, T. Rockta¨schel, S. Riedel, and D. Kiela, “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020,

pp. 9459–9474.

1. S. Kumar, A. Khan, I. A. Banday, M. Gada, and V. V. Shanbhag, “Over- coming LLM Challenges using RAG-Driven Precision in Coffee Leaf Disease Remediation,” in *Proceedings of the 2024 IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2024, pp. 1123–1130.
2. B. Peng, Y. Zhu, Y. Liu, X. Bo, H. Shi, C. Hong, Y. Zhang, and

S. Tang, “Graph Retrieval-Augmented Generation: A Survey,” *arXiv preprint arXiv:2402.23456*, 2024.

1. R. Zhang, H. Du, Y. Liu, D. Niyato, J. Kang, S. Sun, X. Shen, and H.
2. Poor, “Interactive AI with Retrieval-Augmented Generation for Next Generation Networking,” in *Proceedings of the 2024 IEEE International Conference on Communications (ICC)*, 2024, pp. 3456–3463.
3. W. Xu, M. J. Dainoff, L. Ge, and Z. Gao, “Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI,” in *Proceedings of the 2023 ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2023, pp. 1230–1242.
4. S. Jacobs and S. Jaschke, “Leveraging Lecture Content for Feedback Generation with RAG,” in *Proceedings of the 2023 European Conference on Artificial Intelligence (ECAI)*, 2023, pp. 789–798.
5. Y. Huang and J. Huang, “A Survey on Retrieval-Augmented Text Gen- eration for Large Language Models,” in *Journal of Artificial Intelligence Research*, vol. 75, pp. 223–246, 2024.
6. S. Barnett, S. Kurniawan, and S. Thudumu, “Seven Failure Points When Engineering a Retrieval-Augmented Generation System,” in *Proceedings of the 2024 International Conference on Machine Learning (ICML)*, 2024, pp. 1234–1243.
7. P. Chauhan, R. K. Sahani, S. Datta, A. Qadir, M. Raj, and M. M. Ali, “Evaluating Top-k RAG-based Approach for Game Review Generation,” in *Proceedings of the 2024 Conference on Digital Entertainment (CDE)*, 2024, pp. 567–576.
8. A. De Leo, D. Bloxsome, and S. Bayes, “Approaches to Clinical Guideline Development in Healthcare: A Scoping Review and Document Analysis,” in *Healthcare Management Review*, vol. 35, no. 2, pp. 101–115, 2023.
9. S. Inunganbi, “A Systematic Review on Handwritten Document Analysis and Recognition,” in *Proceedings of the 2024 IEEE Conference on Com- puter Vision and Pattern Recognition (CVPR)*, 2024, pp. 2345–2356.
10. O. Weller, B. Van Durme, D. Lawrie, A. Paranjape, Y. Zhang, and J. Hessel, “Promptriever: Instruction-Trained Retrievers Can Be Prompted Like Language Models,” in *Proceedings of the 2024 Conference on*

*Empirical Methods in Natural Language Processing (EMNLP)*, 2024,

pp. 567–576.

1. C. Ravuru, S. S. Sakhinana, and V. Runkana, “Agentic Retrieval- Augmented Generation for Time Series Analysis,” in *Proceedings of the 2024 International Conference on Data Mining (ICDM)*, 2024, pp. 234–243.
2. D. Fleischer, M. Berchansky, M. Wasserblat, and P. Izsak, “RAG Foundry: A Framework for Enhancing LLMs for Retrieval-Augmented Generation,” in *Proceedings of the 2024 International Conference on Computational Linguistics (COLING)*, 2024, pp. 789–798.