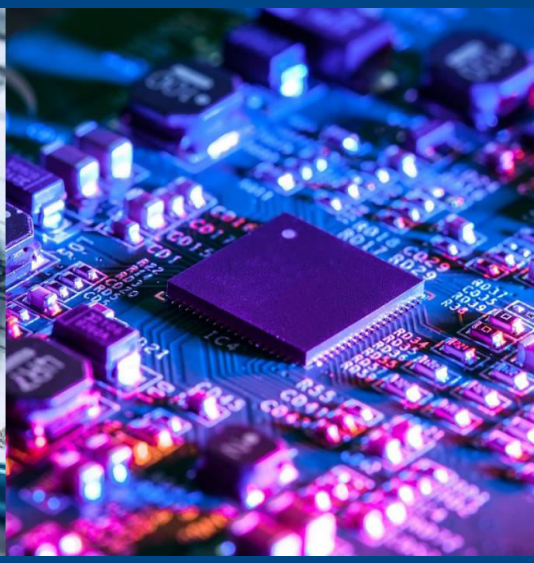
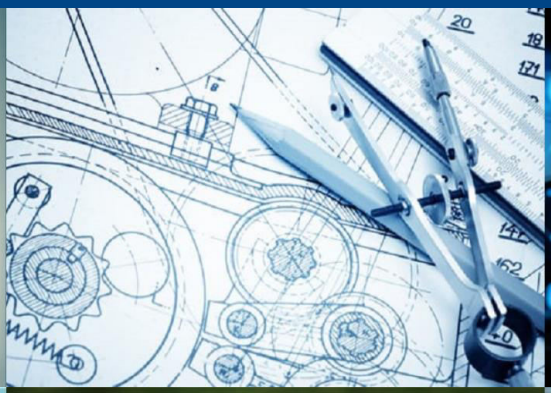
ISSN: 2582-7219

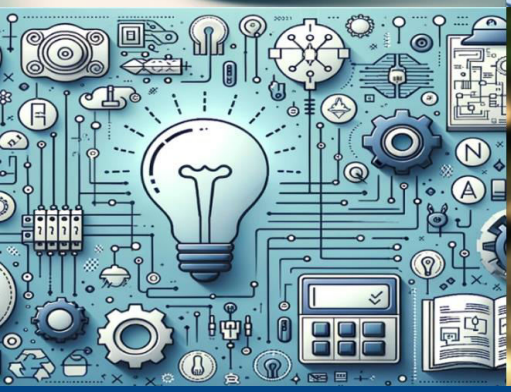




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 3, March 2026**

© 2026 IJMRSET | Volume 9, Issue 3, March 2026| DOI: 10.15680/IJMRSET.2026.0903253

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Campus Hive: Intelligent College Information Assistant using Retrieval-Augmented Generation

S.Lakshmi Priya[1], S.Naveen Kumar[2], A.Vishnulakshmi[3], K.Uthradevi[4]

Assistant Professor, Dept. of Artificial Intelligence and Data Science, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry, India[1]

UG Student, Dept. of Artificial Intelligence and Data Science, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry, India[2,3,4]

ABSTRACT: Educational institutions deal with a lot of documents, emails and files over the place. It is very hard for students and teachers to find what they need. This is where CampusHive comes in. CampusHive is a smart college info assistant. It uses a technique called Retrieval-Augmented Generation. This technique helps CampusHive give answers. CampusHive can search using keywords. Also understand what the words mean. It was built using Amazon Web Services. This makes it very reliable and able to handle a lot of users. The good thing about CampusHive is that it only gives answers based on the information it has. This makes it more accurate than search engines.

KEYWORDS: RAG, Chatbot, Hybrid Search, AWS, NLP

I. INTRODUCTION

Regulations for academic affairs, admission rules, syllabi for courses, examination guidelines, administrative circulars, etc., are some of the many types of information being created and managed within educational institutions. These pieces of information are scattered across various platforms, mostly within the websites and repositories of the respective institutions. These are mostly in unstructured formats, for example, PDF and DOCX files. It is a challenge for the students, teaching staff, and administrative personnel to access the relevant information quickly and effectively due to this scattered pattern of storing the information. As a result, they have to rely on the administrative personnel for day-to-day inquiries, resulting in a lack of efficiency. The existing search systems using keywords are found to be inefficient for searching the required information since they mostly return irrelevant results and are unable to comprehend the meaning of the keywords.

In addition, through the provision of immediate and automated response, such systems, which make use of chatbots, have in the recent past received attention for their ability to improve user interaction and access in educational environments, as discussed in [2], [12]. However, most of the existing chatbots lack domain knowledge integration, and their response is not properly verified, which makes them less reliable. A system that is able to efficiently handle both structured and unstructured data while maintaining response accuracy is also needed.

In order to address these limitations, this paper proposes an intelligent college information assistant named CampusHive that leverages advanced information retrieval and natural language processing techniques. The proposed system will utilize a hybrid search approach that incorporates BM25 keyword search algorithms [7] and vector similarity search algorithms using dense embeddings [4], [6]. Moreover, the proposed system will leverage the Retrieval-Augmented Generation (RAG) framework to improve the accuracy of generated answers by grounding them on relevant document content. This reduces hallucinations in generated answers. The proposed system will also leverage transformer-based language models to improve its ability to comprehend complex user queries and generate context-aware answers.

Moreover, the system utilizes cloud-based data pipeline architectures for scalable storage and processing of institutional documents [5], [10], [17]. With the integration of these technologies, CampusHive is a scalable, efficient, and reliable solution for real-time information access in educational institutions, enhancing user experience and minimizing the need for manual support systems. Furthermore, the proposed system focuses on transparency through source-based

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

answers, increasing user trust and usability. This helps in ensuring the answers are always aligned with the data, making it a feasible solution for real-world applications.

II. LITERATURE SURVEY

Recent developments in Natural Language Processing and Information Retrieval fields have helped create intelligent conversational systems. The document Retrieval-Augmented Generation approach which Lewis et al. [1] created combines document retrieval with language generation to improve factual accuracy. RAG enables generative models to produce responses which depend on knowledge from external sources thus decreasing hallucination instances while increasing model performance. Subsequent research has demonstrated that RAG-based systems are particularly effective in knowledge-intensive applications where accurate and context-aware responses are essential [13], [14].

Chatbot technologies have found extensive research applications in education because they enhance user accessibility while offering immediate assistance and promoting student interaction. Okonkwo [2] provided a complete overview of educational chatbot uses which show their capacity to provide tailored support while decreasing administrative tasks. Winkler and Söllner [12] studied how chatbots function in academic settings and discovered that these systems enhance student learning and classroom participation. Early chatbot systems used rule-based systems and basic machine learning methods which limited their capacity to process advanced questions and handle evolving information. More recent studies demonstrate that RAG-based educational chatbots develop higher response accuracy when researchers combine retrieval mechanisms with generative models according to Swacha et al. [3].

Hybrid search methods now provide an effective solution which can process both structured data and unstructured data. The traditional probabilistic retrieval model BM25 [7] enables efficient keyword-based search matching which search engines use because of its effective and straightforward design. The system faces difficulties because it cannot understand semantic meaning. The word embedding method developed by Mikolov et al. [4] allows textual content to achieve semantic understanding which enables systems to find similar content based on its context. The development of large-scale similarity search methods [6] improves vector retrieval performance by allowing quick searches through high-dimensional embedding spaces. The hybrid search method which combines BM25 with vector similarity search shows better retrieval performance because it uses both lexical and semantic information.

Deep learning model development has brought major advancements to conversational AI systems through its progress in retrieval methods. Transformer architectures which Vaswani et al. [8] introduced changed the field of natural language processing because they allowed text models to understand dependencies between remote text segments. BERT [9] and GPT-based architectures [11] both present strong abilities to understand context while creating responses which sound like human speech. The advancements in language understanding neural architectures which appeared in earlier research [16] functioned as the basis for these developments. Modern chatbot systems now rely on these models as fundamental elements which enable users to interact in a more natural way.

The current study shows that cloud-based architecture has become essential for developing data-driven applications which need both scalability and operational efficiency. The research shows that AWS services which include S3 and Lambda and Glue enable users to build serverless data pipelines and manage extensive data processing operations At present these methods provide effective solutions for data storage and processing and data retrieval which can grow with system demands and maintain system performance. These methods create an effective system which stores data and processes it and retrieves information because they enable system growth and maintain system performance, which makes them appropriate for educational chatbots that require real-time artificial intelligence.

Current systems face multiple obstacles that remain unresolved despite new technological achievements. The majority of chatbot systems operate as black boxes because they fail to show users which sources produced their system-generated responses, which leads to decreased user confidence in their system. The educational sector faces difficulties in managing unstructured institutional data because it requires accurate solutions to handle data from its educational environments. The requirement exists for systems which can manage multiple retrieval methods while keeping their ability to scale and perform. The organization needs to establish an integrated system which unites hybrid search capabilities with advanced language models and domain-specific document references. The CampusHive system solves these problems through its RAG-based system, which uses hybrid retrieval methods to deliver precise and transparent

© 2026 IJMRSET | Volume 9, Issue 3, March 2026| DOI: 10.15680/IJMRSET.2026.0903253

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

information, that educational institutions can use to access information at scale. The system establishes deployment requirements through its focus on three main operational aspects which include building systems that can scale, achieving immediate system response times, and creating systems which easily connect with current institutional systems. The system needs both theoretical development and practical implementation methods to achieve maximum academic

III. METHODOLOGY

The CampusHive system functions as a multi-step system which connects three different processes of document handling and information retrieval and answer production to deliver precise answers which consider both context and special requirements. The system uses unstructured institutional documents to create a knowledge base which allows users to search through documents by using retrieval methods and language models.

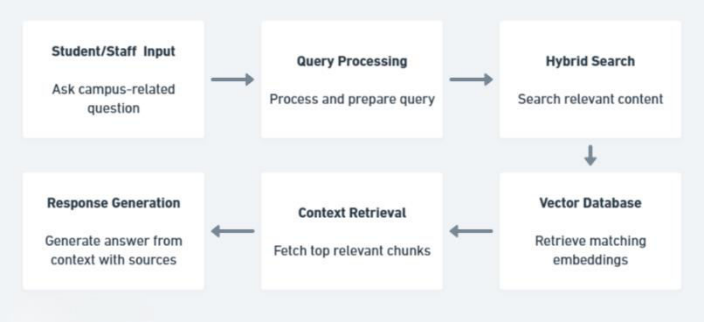


Figure 1: System Workflow

A. Document Collection and Cloud Storage

The system starts its operation by collecting institutional documents which include academic policies and syllabus documents and circulars and administrative guidelines. The documents get uploaded through an administrative interface, which uses AWS S3 for storage because it offers both scalable and reliable cloud storage capabilities. The system controls all data management from a single location, which enables users to access documents needed for their work.

B. Data Preprocessing and Chunking

The documents obtained are analysed to help get the textual content of the formats like PDF, DOCX and TXT. Noise is removed by cleaning the extracted text that contains special characters and wasteful formatting. The cleared material is thereafter separated into smaller logical parts (chunks) in order to maintain contextual meaning. A chunk is linked to metadata, e.g., information about document source and section, so that it is traceable in the retrieval.

C. Embedding and Vector Storage

In order to facilitate semantic processing, every text fragment is transformed into a representation in the form of a vector with the help of embedding methods [4]. These embeddings induce the contextual meaning of the material and permit similarity-based retrieval. The resulting vectors are saved in Weaviate vector database, an efficient indexing and fast retrieval of high-dimensional data. Weaviate vector database is a scalable vector search DB and is supportive of real-time inquiries of huge datasets.

D. Hybrid Retrieval Mechanism

The system operates on hybrid system of retrieval which entails utilization of semantic search and key word search. Accurate keyword matching is carried out on a BM25-based retrieval [7] with the assistance of the most semantically relevant content being retrieved courtesy of the embeddings implemented in Weaviate [4]. The two methods are also combined and prioritized to increase the accuracy of retrieval to ensure that the word relevance is considered as well as the contextual relevance.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

E. Query Processing and Context Formation

After a user makes a query there will be a pre-processing process which will assist in standardization of input and removes noise. The hybrid retrieval mechanism is then applied to the processed query so as to identify the relevant document chunks. The top ranked results are selected and summed up as a contextual input into the model that is generating the response to be utilised so that only the most relevant information is computed.

F. Response Generation using RAG

It is built upon a Retrieval-Augmented Generation (RAG) method and entails retrieved document fragments in the form of a context to produce the responses [1]. A Large language model considers the query and the context obtained and offers the correct and fluent answers. This will ensure that the responses are founded on the information that is retrieved hence reducing hallucination as well as improving reliability.

G. Response Delivery and User Interaction

The resulting response is formulated in a chat bot of a conversation. Also included with the answer in the system are source references and text snippets to increase the transparency of the system. It also displays the confidence score to demonstrate the reliability of a response that enhances the usability and the trust that the user has.

H. Logging, Monitoring, and AWS Integration

All of the user interactions, such as queries, responses and timestamps are logged and stored in S3 to monitor and analyze. AWS Lambda is applied to processing tasks in the backend in a serverless way, which allows executing system operations efficiently and in a scalable manner. AWS Glue would be applicable in data processing and pipeline management. This cloud-based integration makes it highly available, has a high level of scalability and can also manage large datasets efficiently [5].

System Workflow Description

Figure. 1 above presents the workflow of working with user queries and providing answers in the CampusHive system. The process starts with the student or staff input, whereby the user will make queries regarding the campus using the chatbot interface. This is fed to the query processing phase where query gets cleansed and ready to be analyzed.
The submitted query is subsequently sent to the hybrid search element where an information is extracted based on a mixture of semantic search and keyword based search. This step connects with the database of vectors, into which document embeddings are stored and finds corresponding results based on similarity.

The retrieved information is forwarded to the context retrieval phase where the most pertinent document segments are chosen and arranged to compose lucrative context. The response generation component then uses this context to generate an answer based on the information that has been retrieved and using supporting sources.

Lastly, the response that will be generated is relayed to the user in the chatbot display. The workflow shows the consecutive interplay of query processing, retrieval mechanisms, and response generation, which makes the hybrid search and integration of the vector database particularly valuable in the generation of correct and context-sensitive answers.

IV. RESULTS & DISCUSSION

The CampusHive system was introduced in order to determine how well the system can retrieve and produce relevant responses of institutional documents. The system combines the use of hybrid methods of retrieval and Retrieval-Augmented Generation (RA) model to give context-relevant and accurate responses. System retrieval, accurateness of response, and scalability are examined as the performance variables of the system.

A. System Performance

The hybrid search mechanism is effective to process user queries and retrieve the relevant document segments in the system. Relevance of the retrieved results is enhanced by the use of BM25-based keyword [7] and vector similarity search [4], [6]. Semantic matching with embeddings stored in the Weaviate vector database can be efficiently performed, that is, in response to natural language queries that extend beyond matching with exact keys.

© 2026 IJMRSET | Volume 9, Issue 3, March 2026| DOI: 10.15680/IJMRSET.2026.0903253

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Response generation element generates responses with a strict RAG-based response selection methodology [1] as being based on retrieved context, such that the responses are correct and founded in institutional data. The system also gives source references and the created answers, enhancing the transparency and confidence by the user.

B. Retrieval Quality and Accuracy

The hybrid retrieval method has a great contribution towards the accuracy of the system. The retrieval mechanism associated with the key word retrieval is used to provide accurate matching to the structured query use, whereas the search mechanism of the vector similarity is used to locate the semantic links to the complex and natural language query [4], [7]. Such a combination minimizes the irrelevant results and enhances the quality of retrieved information.

The proposed approach exhibits a superior performance compared to single method retrieval systems in terms of information processing ambiguous queries and retrieval of contextually relevant document fragments. The chunk-based retrieval also enhances accuracy because it involves the retrieval of only relevant parts of the documents.

C. Comparative Analysis

In order to compare the efficiency of the offered system, the comparison is made with the widely applied methods like key-word-based retrieval systems and generative chatbot system. **Table 1** shows the results of comparison.

Table 1. Comparison of CampusHive with Existing Systems

Aspect	Keyword-Based (BM25)	Generative Chatbot (LLM)	CampusHive (Hybrid + RAG)
Retrieval	Keyword match	No retrieval	Hybrid (BM25 + Vector)
Understanding	Exact match	Contextual only	Semantic + Context
Accuracy	Medium	Low	High
Hallucination	None	High	Very Low
Transparency	No	No	Yes
Scalability	Moderate	Moderate	High

D. Impact of Cloud and System Integration

The flexibility and resiliency of the system is improved due to the integration of AWS cloud services. AWS S3 is effective in terms of managing large-scale data sets through document storage and serverless computing which helps to handle multiple queries in real time. The pipelines of data that are stored in the clouds enhance the performance and scalability of the systems [5], [10], [17].

Moreover, Weaviate database with the use of its vector database also allows efficient and fast similarity search within huge amounts of embedded data. This integration of cloud infrastructure and vector database will make the performance of the system and the response time much better.

E. Discussion

The findings indicate that hybrid retrieval along with RAG are significant to enhance a chatbot system performance on an educational setting. The proposed system is capable of comprehending the lexical as well as semantic features of user queries unlike the traditional systems, which are based on the use of key words. It generates more credible and fact-based responses as compared to the standalone generative chatbots [1].

The system performance is however a factor of the quality of the input documents as well as the effectiveness of the retrieval process. The possibilities of improvement in future can be based on the optimization of retrieval ranking, integration of personalization, and multilingual queries to increase the system usability and performance further.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION

CampusHive showcases how well the idea of Retrieval-Augmented Generation (RAG) can be implemented to create a smart and trustworthy college information assistant. The system also manages to overcome the difficulties of accessing unstructured institutional data integrating hybrid search methods with a large language model to achieve precise and context-sensitive answers. The system will greatly minimise hallucinations and enhance credibility by making sure that all responses are well-founded on the content of retrieved documents. In addition, hybrid search (BM25 and vector similarity) enhances the accuracy of the retrieval, whereas AWS cloud infrastructure is used to ensure scalability, effective data storing, and stable work of the system. The fact that source attribution scores and confidence scores are included also enhance the transparency and confidence of the user in the system.

In general, CampusHive is a viable and scalable solution to enhancing the accessibility of information in learning institutions. A complete deployment of the cloud, integration of customized response mechanisms, support at other languages, and additional optimization of the retrieval performance are all to be achieved in the future to improve the user experience.

REFERENCES

[1] P. Lewis, E. Perez, A. Piktus, et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *NeurIPS*, 2020.
[2] C. W. Okonkwo, "Chatbots Applications in Education: A Systematic Review," *Computers and Education: Artificial Intelligence*, 2021.
[3] J. Swacha and M. Gracel, "RAG Chatbots for Education: A Survey of Applications," *Applied Sciences*, 2025.
[4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *ICLR*, 2013.
[5] S. O. Amelia and F. Al-Fulan, "Designing Serverless Data Lakes with Amazon S3, AWS Glue, and AWS Lambda for Scalable Analytics Pipelines," *ResearchGate Publication*, 2025.
[6] J. Johnson, M. Douze, and H. Jégou, "Billion-Scale Similarity Search with GPUs," *IEEE Transactions on Big Data*, 2019.
[7] S. Robertson and H. Zaragoza, "The Probabilistic Relevance Framework: BM25 and Beyond," *Foundations and Trends in Information Retrieval*, 2009.
[8] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is All You Need," *NeurIPS*, 2017.
[9] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *NAACL*, 2019.
[10] N. K. Saxena, "Constructing a Scalable Data Pipeline Using AWS Glue and AWS Lambda," *International Journal of Advanced Research in Engineering and Management*, 2025.
[11] A. Radford, J. Wu, R. Child, et al., "Language Models are Unsupervised Multitask Learners," *OpenAI Technical Report*, 2019.
[12] M. Winkler and J. Söllner, "Unleashing the Potential of Chatbots in Education: A State-of-the-Art Analysis," *Academy of Management Proceedings*, 2018.
[13] D. Cooper, "Using Retrieval-Augmented Generation Chatbots to Support Learning," *Journal of Chemical Education*, 2024.
[14] M. Thway, A. Htun, and K. Zin, "A Retrieval-Augmented Generation Chatbot for Higher Education Learning Support," *arXiv*, 2024.
[15] N. K. Saxena, "Constructing a Scalable Data Pipeline Using AWS Glue and AWS Lambda," *International Journal of Advanced Research in Engineering and Management*, 2025.
[16] G. Mesnil, Y. He, L. Deng, and Y. Bengio, "Investigation of Recurrent Neural Architectures for Spoken Language Understanding," *INTERSPEECH*, 2015.
[17] "Building Scalable Serverless Data Pipelines on AWS Using S3, Lambda, and Glue," *International Journal of Scientific Research and Publication*, 2025.
[18] S. Eeti and P. Vashishtha, "Optimizing Data Pipelines in AWS: Best Practices and Techniques," *IJCRT*, 2023.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com