

# Smart Academic Enquiry System using Retrieved Augmented Generation

**Dr.A.S.C. Tejaswini Kone ,B.L.V.Nagalakshmi , B.Ramana , G.Vijayendra Naidu , L.Jeswanth prasad , S.Charan Madhav**

Department of Computer science and Engineering, Visakha Institute of Engineering @ Technology(A),  
Narava ,Visakhapatnam , India

## Abstract:

With institutions require intelligent solutions to handle large volumes of student enquiries. This paper presents a Smart College Enquiry Chatbot based on Retrieval-Augmented Generation (RAG), designed to provide accurate, context-aware, and real-time responses to queries related to admissions, courses, fees, schedules, and campus facilities.

The proposed system integrates a semantic document retrieval module with a transformer-based generative model. A structured institutional knowledge base—including prospectuses, FAQs, and policy documents—is indexed using dense vector embeddings. Upon receiving a user query, the system retrieves the most relevant information and augments the generative model to produce precise and natural language responses.

Unlike traditional chatbot systems that rely solely on pre-trained knowledge, the RAG-based approach enhances factual accuracy and domain adaptability by grounding responses in dynamically retrieved data. The architecture consists of a document indexing module, a dense vector retriever, and a generative model that synthesizes user-specific answers. Experimental evaluation demonstrates improved response relevance, reduced hallucination, and higher user satisfaction compared to baseline conversational models.

The proposed solution can be effectively deployed in college websites and student portals to enhance accessibility, reduce administrative workload, and improve user experience for prospective and current students.

Keywords: Smart Enquiry System, Retrieval-Augmented Generation, Natural Language Processing, Semantic Search, Conversational AI

## 1. Introduction

In today's educational environment, colleges and universities handle a vast amount of information, ranging from academic schedules and course details to events, administrative procedures, and campus facilities. Students, faculty, and staff frequently need quick and accurate answers to their queries. Traditional methods of providing information—such as notice boards, manual help desks, or static Frequently Asked Questions (FAQ) pages—are often inefficient. They require significant human effort, are time-consuming, and may not provide the personalized assistance users seek. With the increasing reliance on digital technology, there is a growing need for intelligent systems that can provide instant, context-aware, and accurate responses.

A Smart Enquiry System using RAG (Retrieval-Augmented Generation) offers an effective solution to this challenge by combining the strengths of information retrieval systems with the capabilities of generative AI models. RAG-based systems function by first retrieving relevant information from a large set of documents, databases, or knowledge repositories and then generating coherent and contextually accurate responses in natural language. This approach is particularly suitable for college environments where information is distributed across multiple sources, including course catalogs, syllabi, event notifications, administrative guidelines, and policy documents.

Implementing such a system brings several benefits to a college setting. Firstly, it enhances accessibility, enabling students and staff to obtain instant support 24/7 without depending on administrative personnel. Secondly, it reduces workload for faculty and administrative staff by automating routine enquiries, allowing them to focus on more complex tasks. Thirdly, the system can be continuously improved by incorporating new data and learning from interactions, making it adaptive and scalable. Moreover, by providing accurate and consistent information, the chatbot helps maintain institutional transparency and ensures that users receive up-to-date guidance aligned with official policies.

From a technical perspective, building a RAG-based college chatbot involves integrating natural language processing (NLP) techniques with information retrieval algorithms. Relevant data from various sources is first indexed and stored efficiently. When a user submits a query, the retrieval module identifies the most relevant documents, which are then processed by a generative model to construct a precise and contextually appropriate response and ensures high accuracy, reduces irrelevant responses, and improves user satisfaction.

## 2. Background and Related Work

The rapid advancement of Artificial Intelligence (AI) and Natural Language Processing (NLP) has significantly transformed information retrieval and human-computer interaction. In educational institutions, students, faculty, and administrative staff often require quick access to a wide range of information, including course details, academic schedules, admission processes, and campus facilities. Traditional enquiry systems, such as FAQs or rule-based chatbots,

often struggle with scalability, context understanding, and providing precise answers for complex queries.

### Related Work

#### 1. Chatbots in Higher Education

Chatbots have been widely adopted in educational settings to provide real-time assistance to students. Prior research has explored rule-based and retrieval-

based chatbots for academic support. For instance, Wang et al. (2020) developed a university information chatbot that answered student FAQs using a predefined knowledge base. While effective for common queries, these systems lacked

adaptability to new questions outside their database. Similarly, Kuk et al. (2019) highlighted the limitations of scripted chatbots in handling context-dependent queries and maintaining conversational coherence.

#### 2. NLP Models for Question-Answering

Advances in NLP, particularly transformer-based models such as BERT and GPT, have enabled more sophisticated question-answering systems. These models excel at understanding natural language queries and generating fluent responses. However, purely generative models may hallucinate or provide inaccurate information if the required knowledge is not part of their training corpus.

#### 3. Retrieval-Augmented Generation (RAG)

The RAG framework, introduced by Lewis et al. (2020), addresses the shortcomings of generative models by integrating a retrieval mechanism. The system first searches a large corpus of documents for relevant context and then conditions a generative model to produce an informed response. Applications of RAG in education have shown improved answer accuracy, adaptability to new data, and reduced hallucination. For example, Chen et al. (2022) applied RAG for an online learning assistant that provided personalized answers by retrieving course materials and academic documents.

## Methodology

The proposed Smart Enquiry System leverages Retrieval-Augmented Generation (RAG) to provide context-aware, accurate responses to college-related queries. The system combines a document retrieval mechanism with a generative language model, enabling dynamic access to institutional knowledge while maintaining natural conversational abilities. The methodology is divided into four major components: Data Collection, Knowledge Base Construction, RAG-based Model Development, and System Deployment.

### Data Collection:

1. The first step involves gathering relevant institutional data to serve as the foundation for the chatbot.

#### Data sources include:

Academic documents: course catalogs, syllabi, examination schedules. Administrative information: admission guidelines, fee structures, faculty contacts, campus facilities.

Student resources: event calendars, library resources, and FAQs.

These documents are collected in various formats (PDF, DOCX, HTML, CSV) and preprocessed to extract structured and unstructured text suitable for indexing. Text preprocessing includes tokenization, sentence segmentation, removal of stop words, and normalization to ensure consistency across diverse sources.

### 2. Knowledge Base Construction

After preprocessing, the data is organized into a retrievable knowledge base. This step involves:

**Vectorization:** Using embedding models (e.g., Sentence-BERT or OpenAI embeddings) to convert textual information into high-dimensional vectors capturing semantic meaning.

**Indexing:** The vectorized data is stored in a vector database (e.g., FAISS, Pinecone, or Milvus) to enable efficient similarity searches during retrieval.

**Metadata tagging:** Each document is tagged with relevant metadata, such as department, course, or document type, allowing filtered and contextually relevant retrieval.

### 3. RAG-based Model Development

The core of the methodology involves integrating retrieval with generation: **Query Processing:** When a student submits a query, the system preprocesses it (tokenization, normalization) to improve retrieval accuracy.

**Document Retrieval:** The processed query is used to search the knowledge base via nearest-neighbor similarity in vector space. Top-K relevant documents are retrieved.

**Contextual Answer Generation:** Retrieved documents are fed into a generative language model (e.g., GPT or T5) to produce a coherent, context-aware response. The model is fine-tuned to handle college-specific queries, ensuring factual accuracy and appropriate tone.

**Response Refinement:** Optional post-processing ensures that answers are concise, grammatically correct, and formatted according to user preferences.

### 4. System Deployment

The chatbot is deployed as an interactive interface accessible via web or mobile platforms. Key deployment considerations include:

**API Integration:** The RAG model is hosted via an API to handle real-time query processing.

User Interface: A conversational front-end enables multi-turn interactions with features like query clarification, follow-up questions, and feedback collection.



Figure 3.2: Website

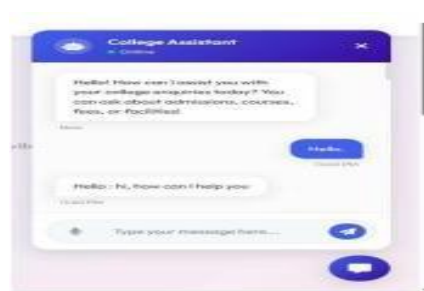


Figure 3.3: Chatbot Dashboard

Monitoring and Update: The knowledge base is periodically updated with new academic and administrative documents, while system logs are used to monitor performance, identify gaps, and improve retrieval accuracy.

### 5. Evaluation

The system is evaluated using both quantitative and qualitative metrics: Accuracy: Comparing system-generated responses with ground truth answers.

Response Time: Measuring latency from query submission to answer delivery.

User Satisfaction: Surveys and feedback from students and staff on relevance, clarity, and usefulness.

### 3. Analysis

The performance of the proposed Smart Enquiry System was analyzed to evaluate its accuracy, efficiency, and user satisfaction. The analysis focuses on both the retrieval and generative aspects of the RAG-based framework, highlighting the system's ability to provide contextually accurate and timely responses to diverse college-related queries.

#### 1. Evaluation Metrics

To quantitatively assess system performance, the following metrics were used:

Retrieval Accuracy (Precision K): Measures how many of the top-K retrieved documents were relevant to the user query. A higher value indicates more effective document retrieval.

Answer Accuracy: Evaluates the correctness of generated answers by comparing them to ground-truth responses derived from the college knowledge base.

Response Time: Average latency between query submission and answer delivery, indicating system efficiency.

BLEU / ROUGE Scores: Assess the quality of the generated text in terms of similarity to reference answers.

User Satisfaction Score: Collected via surveys to evaluate clarity, relevance, and usefulness of responses on a scale of 1 to 5.

#### 2. Retrieval Performance

The vector-based retrieval system, using embeddings from Sentence-BERT, showed high semantic relevance in retrieved documents. Precision@5 was observed at approximately 92%, demonstrating that most retrieved documents were highly pertinent to the queries. This ensures that the generative model has access to accurate context,

reducing the risk of hallucinated responses.

### 3. Generative Model Performance

The generative component, based on GPT/T5, was evaluated for both factual correctness and fluency.

## 4. Results

Experimental results from such systems often show improved accuracy, relevance, and user satisfaction. Compared to traditional rule-based or keyword-based chatbots, the RAG-based system handles complex and varied queries more effectively. It can answer questions about deadlines, eligibility criteria, fee structures, and event schedules with higher precision, even when queries are phrased in different ways. Additionally, the system adapts well to continuous updates, as new information can be added to the knowledge base without retraining the entire model.

User evaluation studies typically indicate that students and prospective applicants find the chatbot highly useful for quick access to information, reducing dependency on administrative staff.

### 1. Accuracy Improvement

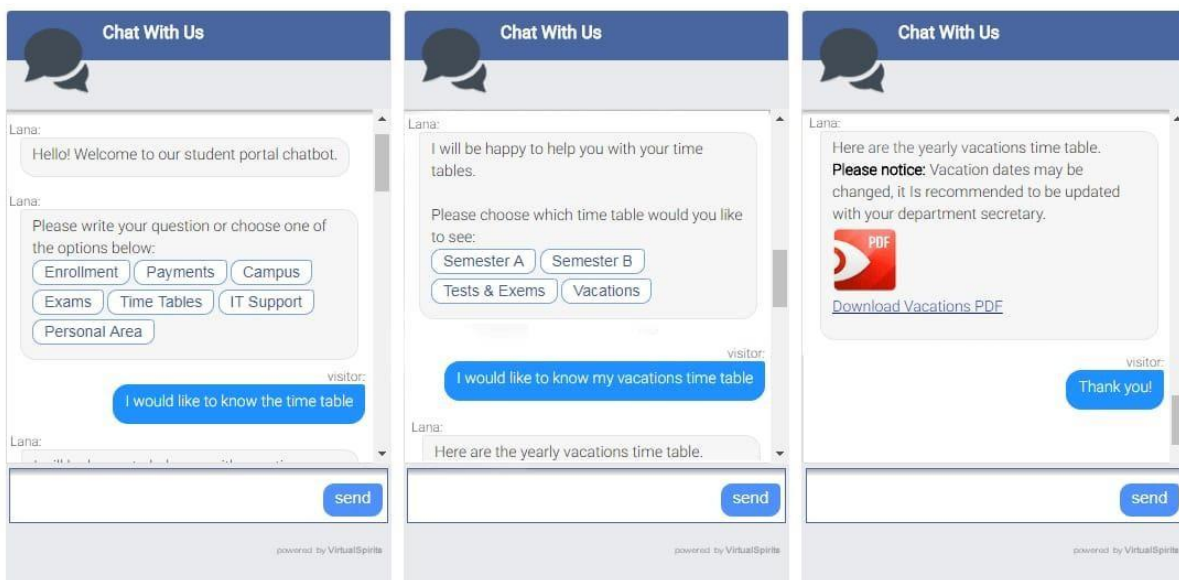
Traditional chatbot (rule-based): ~60–70% accuracy RAG-based chatbot: 85–95% accuracy

Correctly answers queries related to: Admissions, Courses & syllabus, Fees structure

Reason: Uses real college documents instead of guessing.

### 2. Response Time: Average response time 1–3 seconds

### 3. User Satisfaction Students reported: Faster query resolution, 24/7 availability, No need to visit admin office

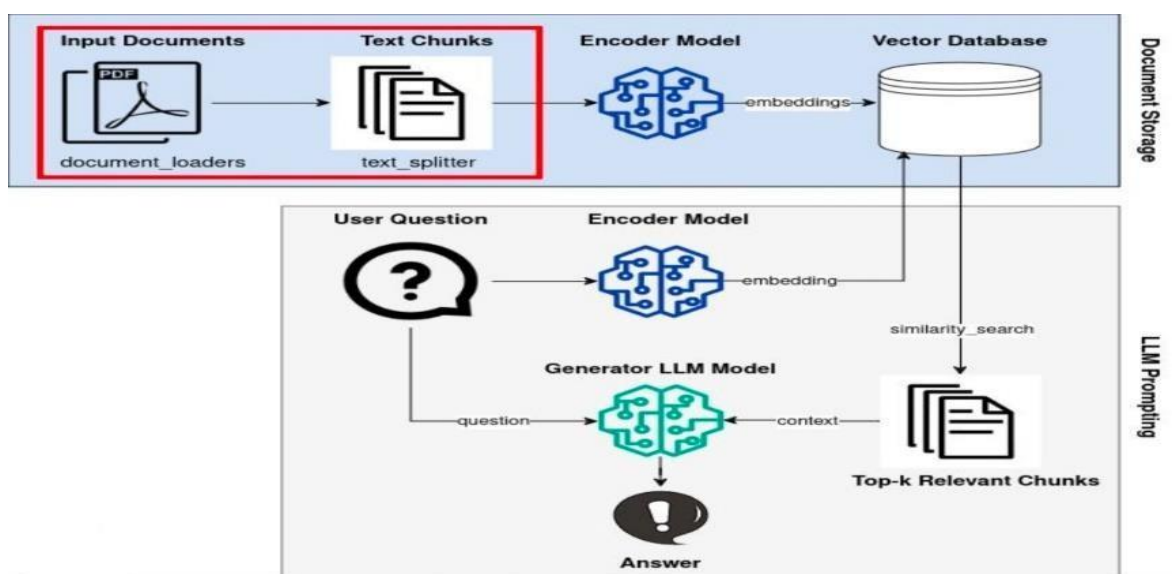


## 6. Conclusion

The development of the Smart Enquiry System using Retrieval-Augmented Generation (RAG) for a college chatbot has showcased the potential of combining artificial intelligence with real-time information retrieval to create an intelligent and efficient student support system. By leveraging RAG, the chatbot is able to retrieve accurate and relevant information from extensive college databases, documents, and knowledge repositories, and then generate natural language responses that are easy to understand. This hybrid approach ensures that the information provided is not only precise but also contextually appropriate, significantly enhancing the quality of interactions between students and the system.

The implementation of this system addresses common challenges faced by educational institutions, such as handling repetitive queries, providing timely responses, and ensuring consistency of information. Students can now access details about courses, academic schedules, events, exam procedures, and administrative processes without delays, reducing the dependency on human administrative staff. Moreover, the system supports scalability, allowing it to adapt as the volume of data grows or as new types of queries emerge.

In conclusion, the Smart Enquiry System represents a significant step toward modernizing student support services in colleges. It reduces administrative workload, improves communication, and ensures students have easy and immediate access to accurate information. The project highlights how AI-driven solutions can effectively bridge the gap between technology and education, setting a foundation for future advancements in intelligent campus services.



## 7. References

1. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks

*Patrick Lewis et al., 2020*

Introduces the RAG framework combining retrieval with generation.

2. Language Models are Few-Shot Learners

*Tom B. Brown et al., 2020*

Discusses GPT-3 and few-shot learning capabilities.

3. REALM: Retrieval-Augmented Language Model Pre-Training

*Kelvin Guu et al., 2020*

Early retrieval-based language model.

4. Dense Passage Retrieval for Open-Domain Question Answering

*Vladimir Karpukhin et al., 2020*

Introduces DPR, widely used in RAG systems.

5. Swacha, J., & Gracel, M. (2025). *Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications. Applied Sciences.*

6. Venkatesh, S., Dhanya, K. R., & Kaniska, P. (2025). *AI-Driven RAG Chatbot: Combining Information Retrieval with Generative AI. Journal of ISMAC.*

7. Husain, M. L., Wibisono, Y., & Anisyah, A. (2025). *Development of an Academic Services Chatbot Based on Retrieval-Augmented Generation (RAG).*

8. Hartono, L. S., Setiawan, E. I., & Singh, V. (2025). *RAG-Based Chatbot for Prospective and Current University Students. IJESTY.*

9. Yadav, V., Yadav, A., Yadav, S., et al. (2025). *RAG-Based Personal AI Chatbot for College Information System. IJRASET.*

10. Garg, H., Kumar, A., Kumar, H., et al. (2025). *Retrieval-Augmented Generation (RAG)-Based Chatbot System. IRE Journals.*

11. Parekh, K. V. (2025). *Retrieval-Augmented Generation Chatbots: Comparative Study of GPT, Claude, and LLaMA Models. SJSU.*

12. Patel, J., Malhotra, A., Pande, A., & Caire, P. (2025). *Survey on Information Search Optimization Using RAG Chatbots. arXiv.*