



AI-BASED COLLEGE INFORMATION ASSISTANT USING RAG

Vaishnavi Nagpure¹

¹ Dattakala Group of Institutions, Faculty of Engineering, Pune, Maharashtra, India, India

Corresponding author: vaishnavinagpure55@gmail.com

Guide / Supervisor: Mrs. Dhage T. S.

ARTICLE ID IJREK-2026-00120	ARTICLE TYPE Research Article	RESEARCH AREA Artificial Intelligence	PUBLICATION Volume 1, Issue Issue 1 (2026)
RECEIVED 19 October 2026	REVISED 20 October 2026	ACCEPTED 20 October 2026	PUBLISHED 20 October 2026

Abstract

Students often need to search across college websites, admission brochures, circulars, academic calendars, fee notices, and departmental documents to find reliable information. This paper proposes an AI-based College Information Assistant using Retrieval-Augmented Generation (RAG), a method that combines information retrieval with language generation. The proposed system accepts a natural-language question, retrieves relevant passages from approved college documents, and uses those passages to generate a concise response with source references. The workflow includes document ingestion, text cleaning, chunking, embedding generation, semantic retrieval, grounded answer generation, and safe fallback handling when evidence is missing or conflicting. The design emphasizes traceability, document freshness, privacy, and avoidance of unsupported claims. A suggested evaluation plan measures retrieval relevance, answer correctness, citation support, abstention quality, response time, and usability. The paper describes the problem, related work, architecture, methodology, implementation options, test scenarios, limitations, and future scope. As no implemented prototype or measured dataset results were supplied, the paper does not claim numerical performance; these must be added after actual testing.

Keywords: Retrieval-Augmented Generation; Large Language Models; College Information System; Semantic Search; Question Answering; Natural Language Processing

I. INTRODUCTION

A. Background / Context Colleges publish important information about admissions, courses, fees, scholarships, examinations, timetables, faculty, and campus facilities through multiple channels. Students and parents may have difficulty locating the right document or confirming whether the information is current. Keyword search can also miss useful material when a question uses different wording from the official source. A conversational assistant can make access easier by allowing users to ask questions in natural language. B. Problem Statement A general-purpose language model may produce fluent but unsupported answers about institution-specific policies. A useful college assistant must retrieve evidence from official sources, present answers clearly, and avoid inventing details when the available documents do not support a response. It should also preserve source links and account for updates or conflicts among college documents. C. Aim and Objectives The aim is to design an AI-based assistant that answers college-related questions using approved documents and RAG. The objectives are to accept natural-language queries; retrieve relevant passages; generate evidence-grounded answers; provide source references; handle uncertain or unanswered queries safely; and define repeatable evaluation criteria. D. Significance / Contribution The proposed design offers a single conversational entry point to college information and a traceable answer workflow. Its intended contribution is the combination of semantic retrieval, source citations, document-version metadata, and safe fallback behaviour. It is a decision-support tool and does not replace official college notices or administrative decisions. E. Paper Organization The following sections review related work, describe the proposed architecture and methodology, present implementation options and test scenarios, define evaluation metrics, and discuss limitations, future scope, and conclusion.

II. LITERATURE REVIEW / RELATED WORK

A. Retrieval-Augmented Generation Lewis et al. introduced Retrieval-Augmented Generation as a framework that combines a retriever over external documents with a

language generator [1]. The approach is relevant to college information because policies and notices can be updated in the knowledge collection without relying solely on the model's internal parameters. However, the quality of the answer still depends on retrieval quality and the generator's use of the evidence. B. Transformer-Based Language Models The Transformer architecture introduced by Vaswani et al. uses attention mechanisms and underpins many modern language models [2]. Such models can interpret user questions and produce readable responses, but fluency does not guarantee factual correctness. In this application, generation should be conditioned on retrieved official passages. C. Semantic Embeddings and Vector Search Sentence-BERT provides sentence embeddings that can be compared for semantic similarity [3]. This supports retrieval when the user's wording differs from the wording of a college document. FAISS is a library for efficient similarity search and clustering of dense vectors and can be used as one option for a local vector index [4]. D. Research Gap A practical institutional assistant must address more than question generation. It needs document provenance, freshness checks, citation support, privacy controls, and an explicit response when evidence is missing. The proposed paper focuses on these operational requirements and a repeatable evaluation plan. The design is proposed and has not been presented as a completed empirical comparison.

III. PROPOSED SYSTEM AND METHODOLOGY

A. System Architecture The system contains a knowledge-source layer, document-ingestion pipeline, embedding and index layer, query-retrieval module, grounded-generation module, and user interface. Approved college documents are parsed and divided into chunks. Each chunk is embedded and stored with source metadata. At query time, relevant passages are retrieved and provided to the language model, which generates an answer with references. Approved College Sources Web pages PDFs Circulars FAQs ↓ Document Processing Text extraction cleaning chunking metadata ↓ Embedding and Vector Index Vector representation source tracking ↓ Question Retrieval Query embedding similarity ranking relevance check ↓ Grounded Answer Generation Answer from retrieved evidence abstain if unsupported ↓ Student Interface Answer source links feedback

Fig. 1. Proposed architecture of the AI-based college information assistant. B. Knowledge Sources and Document Preparation The knowledge base may include official admission notices, prospectuses, fee schedules, course descriptions, scholarship information, examination rules, academic calendars, and student FAQs. Each document should retain its title, URL or file identifier, department, publication date, effective date where available, version, and last-ingested timestamp. Text extraction should preserve headings and relevant table structure. Scanned files may require OCR and manual checks for dates, amounts, and eligibility conditions. C. Chunking and Embedding Cleaned text is divided into manageable chunks that preserve enough context to answer a question. Chunk size and overlap should be selected through testing. An embedding model converts each chunk into a vector representation. The index stores vectors alongside the source document and section or page information so that the system can cite the evidence used. D. Retrieval and Answer Generation For each query, the system computes a query embedding and retrieves the most relevant chunks using a similarity measure. Hybrid keyword and semantic retrieval may improve handling of exact course codes, dates, and named policies. A relevance threshold can prevent weak matches from being passed as evidence. The language model receives the question and retrieved text with instructions to answer only from the evidence, identify uncertainty, and provide citations. E. Working Flow The main stages are: (1) accept and validate the query; (2) search the indexed college documents; (3) rank relevant passages; (4) check whether evidence is sufficient; (5) generate a concise answer with source references; and (6) display a safe fallback or contact guidance when no reliable answer is found. Question → Retrieve → Generate → Answer + Sources

Fig. 2. Query processing and response workflow. F. Suggested Algorithm Receive the user's question and validate input. Convert the query to an embedding and search the indexed document chunks. Rank retrieved passages and apply relevance and access-control checks. If evidence is sufficient, generate an answer using only the selected context. Attach source links or document/page references to the response. If evidence is absent or conflicting, state the limitation and direct the user to an official source. Record only necessary feedback and operational metadata under the institution's privacy policy.

IV. IMPLEMENTATION DETAILS

TABLE I		
Component	Possible implementation	Purpose
User interface	HTML, CSS, JavaScript or React	Question entry, answers, source links
Backend API	Python with FastAPI or Flask	Ingestion, retrieval, access checks
Embeddings	Sentence-Transformers or compatible API	Vectorize text and queries
Vector index	FAISS or managed vector database	Similarity search
Language model	Suitable hosted or local LLM	Generate grounded response
Storage	Relational database and document store	Metadata, versions, feedback

V. RESULTS AND ANALYSIS

TABLE II		
Test ID	Input / condition	Expected behaviour
T1	Question answered directly in an official FAQ	Returns a concise answer with a valid source reference.
T2	Question paraphrases wording in a notice	Semantic retrieval finds the relevant passage.
T3	Answer is absent from the document collection	States that evidence is unavailable; does not invent details.
T4	Two documents contain conflicting dates	Flags the conflict or prioritizes an explicitly current official version.
T5	Outdated notice has been superseded	Uses the current notice or identifies the older source as superseded.
T6	Query requests restricted student data	Enforces authorization and does not expose private data.



Fig. 3. Conceptual design priorities only; bars are not measured model results.

VI. LIMITATIONS AND FUTURE SCOPE

A. Limitations The proposed system has not been empirically evaluated in this paper. Its quality depends on whether the knowledge base is complete, current, and correctly extracted. Retrieval can miss relevant passages, and a language model can still misstate evidence. Scanned PDFs, tables, multilingual documents, and conflicting notices require additional processing. Hosted models may introduce recurring costs and data-handling considerations. The system should not be treated as an official decision-maker. B. Future Scope Future work may add English, Hindi, and Marathi support; integrate approved live notices and academic calendars; test hybrid retrieval and re-ranking; build an administrator dashboard for document updates; add voice-based questions and accessibility improvements; and conduct a user study. The assistant can also collect privacy-conscious feedback to identify unanswered questions and improve the official FAQ collection.

VII. CONCLUSION

This paper proposes an AI-Based College Information Assistant using Retrieval-Augmented Generation to make college information easier to access through natural-language questions. The system retrieves passages from approved institutional documents, uses them to ground generated answers, and provides source references so users can verify important details. The design includes document versioning, relevance checks, privacy controls, and fallback behaviour when evidence is missing or conflicting. Its effectiveness must be established through implementation and testing on representative college questions; no unverified performance claims are made here. Reliable operation depends on current official sources, secure access controls, careful evaluation, and human oversight.

REFERENCES

[1] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Advances in Neural Information Processing Systems*, vol. 33, 2020. [Online]. Available: <https://arxiv.org/abs/2005.11401>.
[2] A. Vaswani et al., "Attention Is All You Need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: <https://arxiv.org/abs/1706.03762>.
[3] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proceedings of EMNLP-IJCNLP*, pp. 3982–3992, 2019, doi: 10.18653/v1/D19-1410. [Online]. Available: <https://aclanthology.org/D19-1410/>.
[4] M. Douze et al., "The Faiss library," *arXiv:2401.08281*, 2024. [Online]. Available: <https://arxiv.org/abs/2401.08281>. Note: This paper follows the compact section structure of the supplied sample. Add the institution name and logo if required by your college. Actual implementation and measured results should be included only after verification.

Publication archival note: Official scholarly contribution published under CC BY 4.0 International License. Permanent record on <https://www.ijrek.org> | Article ID: IJREK-2026-00120