

AI-Integrated Campus Genie

Sohel Tamboli¹, Sanjana Telpure², Rutuja Thite³, Nirmala Shinge⁴

Department of BCA, School of Computational Sciences, Faculty of Science and Technology

JSPM University, Pune, Maharashtra, India

soheltamboli23.ca@jspmuni.ac.in¹, sanjanatelpure23.ca@jspmuni.ac.in², rutujathite23.ca@jspmuni.ac.in³,
nss1.scos@jspmuni.ac.in⁴

Abstract: Educational institutions face an ever-growing volume of administrative queries from students, parents, and staff regarding admissions, fee structures, examination schedules, academic calendars, and policy updates. Traditional methods such as enquiry counters, phone lines, or static websites are inefficient due to limited operating hours, long waiting times, and language barriers. To address these challenges, this paper proposes AI-INTEGRATED CAMPUS GENIE, a language-agnostic, AI-powered chatbot that leverages Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs). The system provides accurate, context-aware, and multilingual responses by intelligently retrieving and processing information from official institutional documents. Key features include document ingestion with FAISS-based vector embeddings, multilingual query translation, session-aware context handling, PII sanitization, and high-speed inference via Groq's LPU architecture.

Keywords: Chatbot; Retrieval-Augmented Generation (RAG); Large Language Model (LLM); FAISS; Multilingual AI; Campus Automation; Natural Language Processing (NLP)

I. INTRODUCTION

In the digital era, students and parents expect instant, accurate information from educational institutions. Administrative offices frequently receive a high volume of repetitive queries about admissions, examination schedules, fee structures, and academic policies. Existing methods—including enquiry counters, institutional websites, telephone support, and Whats App groups—are inadequate due to limited operating hours, long waiting times, unstructured information, and language barriers for non-English speakers.

Rule-based chat bots lack flexibility and cannot understand natural language nuances, while generic AI models often produce hallucinated or inaccurate responses. AI-INTEGRATED CAMPUS GENIE addresses these challenges by providing real-time retrieval from official documents, accurate and context-aware responses, multilingual support for regional languages, and privacy protection for sensitive student data.

II. PROBLEM STATEMENT

Educational institutions encounter several systemic challenges in handling administrative queries:

- High administrative workload due to repetitive and routine queries.
- Delayed responses arising from limited office hours and staff availability.
- Language diversity restricting accessibility for non-English-speaking stakeholders.
- Critical information buried in non-searchable PDF documents or static web pages.
- Absence of a unified system for accurate, instantaneous query resolution.

There is a pressing need for a system that understands natural language queries, retrieves precise information from official documents, provides multilingual support, and reduces manual administrative effort.



III. OBJECTIVES OF THE STUDY

The main objectives of this research are:

- Develop a RAG-based chatbot to intelligently access and retrieve institutional data.
- Support multilingual communication in Hindi, Marathi, Tamil, Telugu, and other regional languages.
- Implement vector-based retrieval using FAISS for fast and accurate semantic search.
- Ensure user privacy by sanitizing sensitive personal data (PII) before processing.
- Provide quick and precise responses sourced exclusively from official institutional documents.
- Reduce administrative workload by automating repetitive query handling.

IV. LITERATURE REVIEW

A. Retrieval-Augmented Generation (Lewis et al., 2020)

RAG combines document retrieval with language model generation, significantly reducing hallucinated responses and enhancing factual accuracy for knowledge-intensive tasks [1].

B. Transformer Architecture (Vaswani et al., 2017)

The transformer model introduced the self-attention mechanism and serves as the foundational architecture for modern LLMs including GPT and LLaMA [2].

C. GPT Models (Brown et al., 2020)

GPT-3 demonstrated strong zero-shot and few-shot learning capabilities, providing robust natural language understanding for contextual response generation [3].

D. FAISS (Johnson et al., 2017)

FAISS (Facebook AI Similarity Search) enables efficient similarity search over large datasets, facilitating fast retrieval of relevant document embeddings at scale [4].

E. Multilingual Translation Systems (Google, 2016)

Google's multilingual neural machine translation supports multiple languages within a single model, facilitating cross-language communication [5].

F. LangChain Framework (2024)

LangChain simplifies RAG pipeline creation by providing integrated tools for embedding generation, document processing, and chain-of-thought prompting [6].

G. Groq LPU Architecture (2024)

Groq's Language Processing Unit (LPU) provides hardware-accelerated LLM inference with low latency, making it suitable for real-time applications [7].

Prior research emphasizes the importance of combining retrieval and generation strategies, incorporating multilingual support, ensuring privacy, and enabling efficient embedding retrieval for accurate conversational AI systems.

V. EXISTING SYSTEM ANALYSIS

Current methods for handling administrative queries remain largely inefficient:

Manual Enquiry Counters: Restricted to office hours; students face long queues and significant time overhead for both staff and visitors.

Telephone Support: Lines are frequently busy, and no permanent documentation of queries or resolutions is maintained.



Institutional Websites: Content is static and difficult to navigate, particularly on mobile devices.

WhatsApp/Group Chats: Responses are unverified, unofficial, and cannot be reliably referenced.

VI. PROPOSED SYSTEM: AI-INTEGRATED CAMPUS GENIE

AI-INTEGRATED CAMPUS GENIE is a web-based, RAG-powered chatbot capable of handling queries in multiple languages. The system workflow proceeds as follows:

1. User submits a query in any supported language.
2. The sanitization module removes sensitive personal data (PII) from the input.
3. The query is translated to English for uniform internal processing.
4. Relevant context is retrieved from the FAISS vector database using semantic similarity search.
5. The LLM generates a grounded response based on the retrieved context.
6. The response is translated back into the user's original language and returned.

VII. SYSTEM ARCHITECTURE

The system is organized into five functional layers:

A. Client Layer

Provides the web interface for query input and user interaction. Supports multiple languages at the presentation level.

B. Application Layer

A Flask server manages all incoming requests, session handling, and response delivery.

C. Processing Engine

Encompasses the privacy filter for PII removal, a translation module for multilingual support, a context retriever for maintaining conversation history, and a prompt builder for constructing LLM input.

D. Data Layer

Hosts the FAISS vector database for embedding storage and retrieval, and document storage for PDFs and other institutional records.

E. AI Layer

Employs LLaMA (via Groq) to generate context-aware, document-grounded responses with sub-3-second latency.

VIII. METHODOLOGY

A. Ingestion Module

Institutional documents (PDFs, CSVs, circulars) are uploaded to the system.

Documents are split into manageable text chunks using a sliding-window approach.

Embeddings are generated using the MiniLM sentence transformer model.

Embeddings are stored in FAISS for fast approximate nearest-neighbor retrieval.

B. Chat Processing Module

PII (phone numbers, emails, student IDs) is detected and removed before any external API call.

User queries are translated to English via the Deep Translator API.

Conversation history is maintained in session state to support follow-up queries.

C. Generation Module

A structured prompt is built from the retrieved context and conversation history.

The LLM (LLaMA via Groq API) generates a factually grounded response.

The response is translated back to the user's language and returned.



IX. TOOLS AND TECHNOLOGIES

A. Software and Frameworks

Python 3.9+, Flask web framework, HTML/CSS/JavaScript for front-end.

B. Libraries

LangChain (RAG pipeline), FAISS (vector search), HuggingFace Transformers (MiniLM embeddings), PyPDFLoader (document parsing).

C. APIs

Groq API for high-speed LLM inference; Google/Deep Translator API for multilingual translation.

X. RESULTS AND EXPECTED OUTCOMES

Based on system design and preliminary evaluation, the following performance targets are established:

- Average response time below 3 seconds per query.
- Retrieved answer accuracy exceeding 90% for in-domain institutional queries.
- Multilingual support for five or more regional languages.
- Measurable reduction in administrative staff workload.
- Improved student and parent satisfaction with information access.

XI. APPLICATIONS

A. College Admission and Registration

Provides real-time guidance on admission criteria, eligibility requirements, required documents, and deadlines. Automates routine admission queries, reducing staff burden and guiding users through registration and fee payment workflows.

B. Examination Scheduling

Responds to queries regarding exam dates, timetables, venues, and procedures. Maintains conversational context to handle multi-turn follow-up questions such as “What is my exam schedule for semester 2?”

C. Library and Resource Management

Provides information about available books, digital resources, and library hours. Assists with book issuance inquiries, due dates, and fines. Supports multilingual access to library services.

D. Public Information and Campus Navigation

Acts as a virtual assistant for campus visitors, handling queries about facilities, events, workshops, seminars, and official notices.

E. Government and Institutional Services

Assists with scholarship and financial aid applications, student ID issuance, hostel accommodations, and transportation. Disseminates policy updates and official circulars in multiple languages.

XII. ADVANTAGES AND LIMITATIONS

A. Advantages

Multilingual Support: Users interact in Hindi, Marathi, Tamil, Telugu, and other regional languages, ensuring inclusivity.

Reduced Hallucination: RAG grounds all responses in official documents, minimizing fabricated information.



Fast Retrieval: FAISS vector search combined with Groq's LPU delivers responses in under 3 seconds.

Privacy Protection: PII is automatically sanitized before any external processing, ensuring data protection compliance.

24/7 Availability: The virtual assistant operates continuously, eliminating delays caused by office hours.

B. Limitations

Computational Overhead: Running LLMs and RAG-based retrieval requires significant CPU/GPU resources, which may challenge institutions with limited hardware.

Document Quality Dependency: Response accuracy depends on the completeness and formatting of uploaded institutional documents.

Translation Inaccuracies: Automatic multilingual translation may introduce minor errors affecting comprehension in non-English responses.

Indexing Requirements: Effective retrieval requires careful chunking, embedding, and indexing; improper indexing may degrade search quality.

XIII. FUTURE SCOPE

- Voice-enabled query support, allowing speech-based interaction with the chatbot.
- Mobile application integration for on-the-go access via iOS and Android platforms.
- Image-based query processing using OCR to extract and respond to text in scanned notices.
- Integration with ERP and student management systems for real-time data access (attendance, fees, records).
- AI-based analytics dashboard for query pattern analysis, institutional insights, and service improvement.

XIV. CONCLUSION

AI-INTEGRATED CAMPUS GENIE demonstrates a robust approach to modernizing student-institution interaction by addressing the limitations of traditional administrative systems. By integrating Retrieval-Augmented Generation with Large Language Models, the system ensures accurate, context-aware, and verifiable responses sourced directly from institutional documents.

Its language-agnostic architecture supports multilingual queries, enabling students and parents from diverse linguistic backgrounds to access critical information without barriers. The PII sanitization module ensures compliance with data protection standards, while FAISS-based vector retrieval and Groq's high-speed inference reduce response times to under three seconds.

Beyond routine query handling, the platform provides a foundation for scalable automation and data-driven institutional insights. Planned enhancements—including voice interaction, mobile integration, image-based processing, ERP connectivity, and AI-powered analytics—will evolve the system into a comprehensive smart campus assistant. AI-INTEGRATED CAMPUS GENIE represents a significant step toward digital transformation in educational institutions, offering a scalable, secure, and inclusive solution for administrative communication and student engagement.

REFERENCES

1. P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459–9474.
2. A. Vaswani et al., "Attention is All You Need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
3. T. B. Brown et al., "Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 1877–1901.
4. J. Johnson, M. Douze, and H. Jégou, "Billion-Scale Similarity Search with GPUs," *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547, 2021.



5. M. Johnson et al., "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation," Transactions of the Association for Computational Linguistics, vol. 5, pp. 339–351, 2017.
6. Harrison Chase, "LangChain," 2022. [Online]. Available: <https://github.com/langchain-ai/langchain>.
7. Groq Inc., "Groq Language Processing Unit (LPU) Architecture," 2024. [Online]. Available: <https://groq.com>.

