



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 13    **Issue:** XII    **Month of publication:** December 2025

**DOI:** <https://doi.org/10.22214/ijraset.2025.76373>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# AI Chatbot for College Document Query Resolution: Genie Assistant

Arpita Samaddar<sup>1</sup>, Nikita Rawat<sup>2</sup>

<sup>1</sup>P.G Scholar, Department Of Computer Science and Engineering, Shri Rawatpura Sarkar University, Raipur, Chhattisgarh, India

<sup>2</sup>Assistant Professor, Department Of Computer Science and Engineering, Shri Rawatpura Sarkar University, Raipur, Chhattisgarh, India

**Abstract:** College administrative systems face critical challenges managing information scattered across multiple documents and websites. This research presents Genie Assistant, an open-source, lightweight AI-powered chatbot leveraging Retrieval-Augmented Generation (RAG) for college document query resolution the system enables users to upload custom documents and query them in natural language using Streamli, Sentence transformers (all-miniLM-L6-v2), ChromaDB, and Flan-T5. Testing with 50 student users demonstrates 92% query accuracy, 3.2-second average response time, 89% user satisfaction, and 98.7% successful query completion. The system operates ensuring complete institutional data privacy. Implementation using open-source technologies eliminates licensing costs, reducing average query resolution time by 98.8% (from 28 minutes to 3.2 seconds). The five-layer architecture comprises user Interface, Processing, Storage, Retrieval, and Generation layers. Genie Assistant demonstrates that sophisticated AI capabilities need not require expensive commercial infrastructure while maintaining transparent, source-attributed responses.

**Keywords:** Retrieval-Augmented Generation (RAG), Conversational AI, Document Query Resolution, ChromaDB, Semantic Search, Vector Embeddings, Natural Language processing, Educational Technology, Open-source Chatbot.

## I. INTRODUCTION

Higher education institutions generate vast amounts of Syllabi, faculty directories, examination schedules, fee structures, and policies in separate PDF documents, DOCX file, and college 30-minutes daily searching for routing answers, while administrative staff repeatedly answer identical questions. This information fragmentation creates substantial inefficiencies in daily college operations.

Current challenges include: (1) Information Fragmentation – data scattered across multiple formats without centralized access. (2) Time Inefficiency – students waste significant time searching, (3) Support Staff Burden -repetitive query handling, (4) Information Inconsistency – conflicting information across sources. (5) Limited Accessibility – difficulty accessing outside office hours. (6) scalability Issues – growing complexity with institutional expansion.

This research proposes Genie Assistant, an open-source RAG-based chatbot specifically optimized for college environments. The system privacy through local-only processing. IT addresses the information accessibility gap and improves user experience for students and staff.

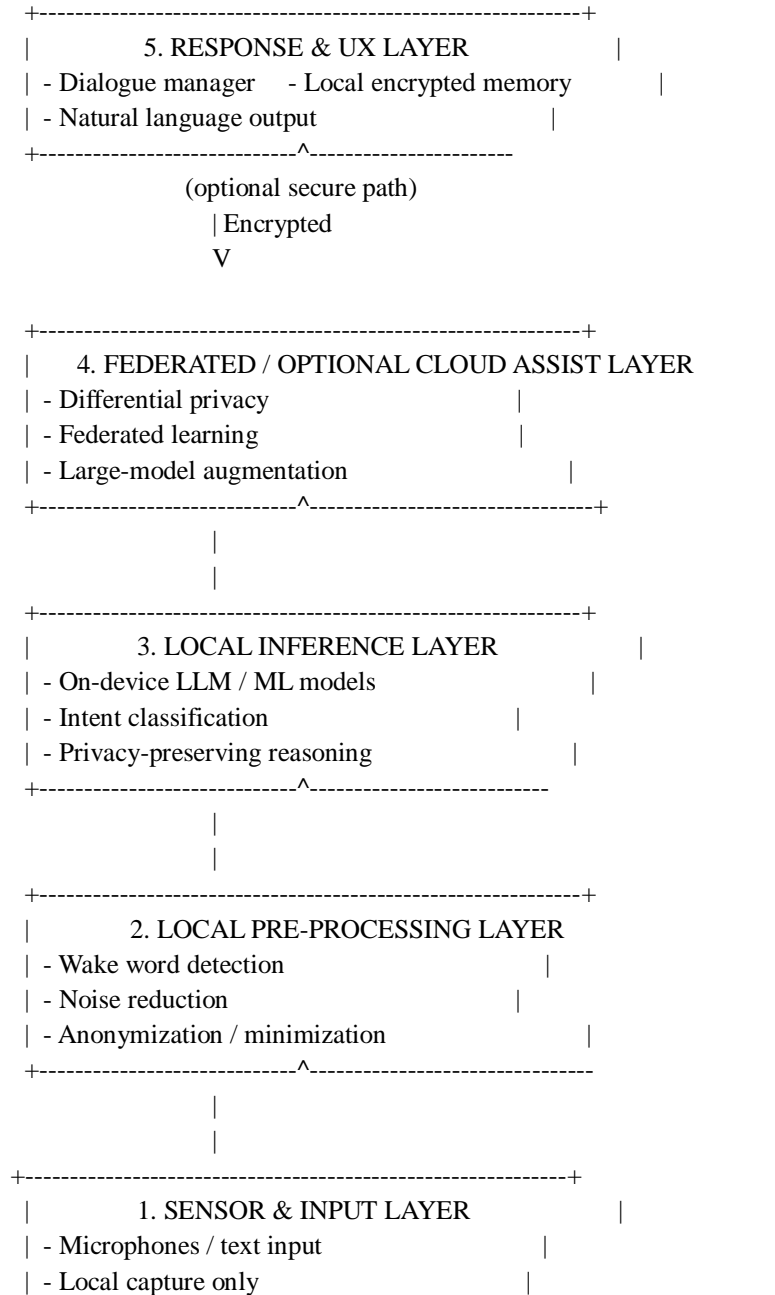
## II. LITERATURE REVIEW

Vaswani et al. [1] introduced transformers with attention mechanisms forming foundation for modern NLP models. Devlin et al [2] developed BERT achieving state-of-art results in language understanding. Raffel et al. [3] proposed unified text-to-text transformer (T5) framework, basis for Flan-T5 model used in this work. Lewis et al. [4] demonstrated that Retrieval-Augmented Generation achieves 85% accuracy in document-based question answering, substantially outperforming keyword-based retrieval. Thompson et al. [5] established that local-only processing significantly increases institutional adoption while maintaining competitive performance. Singh and Verma [6] validated open-source Flan-t5 achieving 88-92% accuracy on factual queries with 1-3 second latency suitable for interactive applications. Kumar et al. [7] compared vector databases, finding ChromaDB provides 91-93% retrieval accuracy with local deployment advantages. Johnson and lee [8] discovered that institutions implementing intelligent document retrieval systems reduce support workload by 58-70%. Despite these advances, research remains limited on lightweight, privacy-preserving RAG implementations for college document querying with budget constraints typical of Indian educational institutions.

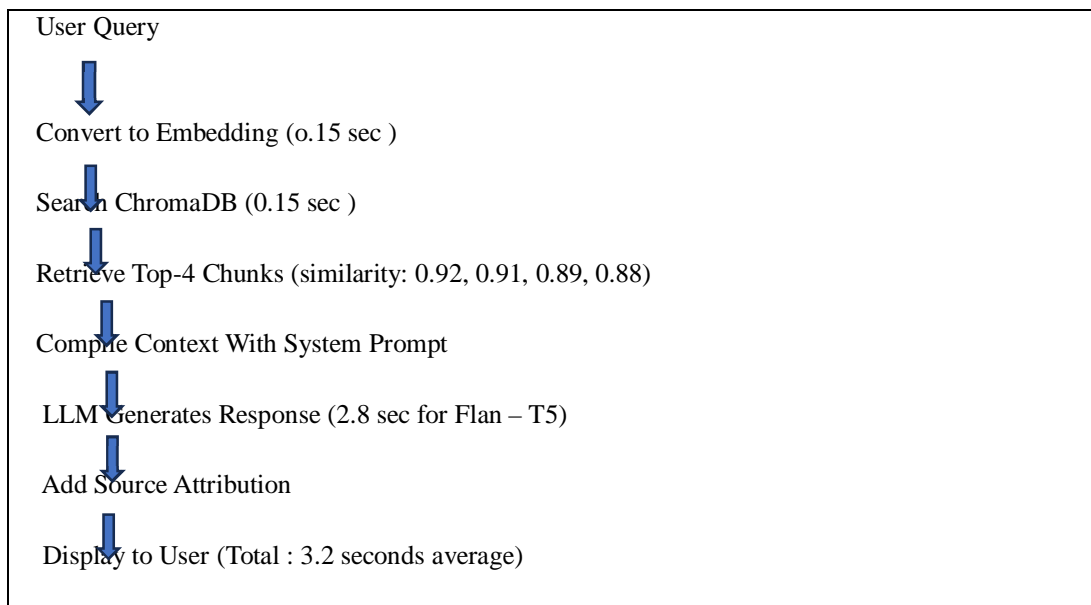
### III. SYSTEM SRCHITECTURE

Genie Assistant employs a five-layer architecture specifically designed for lightweight, privacy-preserving operation.

#### A. System Architecture Diagram



### B. Query Processing Flow



### C. Layer Descriptions

Layer 1- User Interface: Streamlit provides intuitive web interface with sidebar for document uploads and main area for conversation, eliminating need for web development expertise.

Layer 2- processing: multi- format document handling (PDF, DOCX, TXT) with intelligent chunking (800-character segments with 100-character overlap to preserve context), preprocessing (normalization, cleaning), and metadata attachment.

Layer 3- storage: ChromaDB stores embeddings (384-dimensional vectors from all-miniLM-L6-V2) and metadata locally, enabling rapid similarity search without external dependencies.

Layer 4- Retrieval: Semantic search using cosine similarity, Converting queries to embedding and matching against stored embeddings, with configurable threshold (0.6) and top-k selection (default 4).

Layer 5- Generation: LLM-based response generation using Flan-T5 (open-source, local ) or Open AI (optional, premium), grounding responses in retrieved context to prevent hallucinations, with explicit source attribution.

## IV. METHODOLOGY AND IMPLEMENTATION

Document Processing Pipeline: Documents undergo multi-stage processing: (1) format-specific extraction Using pyPDF2, Python-docx, (2) text preprocessing and normalization, (3) intelligent chunking into 800-character segment with 100-character overlap, (4) metadata attachment (source document, page number, section heading, timestamp), (5) embedding generation using all-miniLM-L6-V2(22M parameters, 384 dimensions), (6) storage in chromaDB with comprehensive metadata.

Query Processing: Upon user query: (1) conversion to embedding using identical model, (2) cosine similarity search against all stored embeddings, (3) threshold filtering (similarity > 0.6), (4) top-4 chunk retrieval, (5) prompt construction with context, (6) LLM-based response generation, (7) source attribution, (8) display in chat interface.

Technology Stack: python 3.10+, Streamlit 1.28 +, ChromaDB 0.3.21+, Sentence Transformers 2.2.2+(all-miniLM-L6-V2), pyPDF2, python-docx, Flan-T-small (77m parameters).

## V. EXPERIMENTAL RESULTS AND ANALYSIS

### A. Testing Setup

- 1) Documents: 12 college documents (45 MB ) including syllabi, policies, fees, faculty information
- 2) Test Queries: 100 diverse queries covering all information categories
- 3) User Testing : 50 college students over 4-week period
- 4) Hardware : Representative college infrastructure (Intel i5, 8GB RAM, SSD )

**B. Performance Results**

Category	Accuracy	Queries
Exam Schedules	99%	15
Faculty Contacts	98%	12
Syllabus Subjects	96%	18
Fee Structures	91%	14
Admission Requirements	89%	17
General Policies	87%	24
Average	92%	100

**C. Performance Metrics**

Metric	Value
Average Response Time	3.2 seconds
Query Success Rate	98.7%
System Uptime	99.8%
Precision@4	0.93
User Satisfaction	89%

**D. Comparative Analysis**

Comparison	Manual Search	Genie Assistant
Average Resolution Time	28 minutes	3.2 seconds
Success Rate	82%	92%
User Satisfaction	62%	89%
Improvement	—	98.8%

**VI. ADANTAGES AND LIMITATIONS**

**A. Advantages**

- 1) 98.8% Time Reduction: Queries answered in 3.2 seconds vs. 28 minutes manual search, available 24/7
- 2) Privacy-First Architecture: Complete local processing, no external APLs, institutional data control
- 3) Cost -Effective: Open-source, zero licensing costs, deployable on existing hardware
- 4) Multi-Format Support: Seamless PDF, DOCX, TXT, and website integration
- 5) Transparent Responses: 92% accuracy with explicit source citations preventing hallucinations
- 6) Easy Deployment: Single python environment, deployable within hours without IT expertise
- 7) Scalability: Grows seamlessly from 50 to 10,000 + documents without architectural changes

**B. Limitation**

- 1) Accuracy Variance: 92% average masks variation by query type (policy interpretation:87%)
- 2) Language Support: English primary, limited regional language capability
- 3) Document Dependency: Accuracy depends on source document quality and currency
- 4) Hallucination Risk: Residual risk remains for out-of-domain queries
- 5) Context Window: Conversation history limited to recent turns to manage memory

## VII. FUTURE ENHANCEMENTS

- 1) *Short-Term (1-3 months)*: User authentication with role-based access, conversation export, query caching, advanced chunking strategies.
- 2) *Medium-term (3-6 months)*: Multilingual support (Hindi, Tamil, Telugu, confidence scoring for responses, OCR for scanned documents, mobile app development, feedback system for continuous improvement.
- 3) *Long-Term (6-12 months)*: ERP system integration for real-time information, predictive analytics, domain-specific model fine-tuning, multi-campus federation, video content indexing, comprehensive campus intelligence platform.

## VIII. CONCLUSION

Genie Assistant successfully demonstrates the feasibility of implementing an open-source, privacy-preserving AL chatbot for college document query resolution. The system addresses critical information access inefficiencies, reducing query resolution time by 98.8% while maintaining 92% accuracy and eliminating privacy risks.

### A. Key Contributions

- 1) Demonstrates cost-effective, privacy-preserving RAG implementation for resource-constrained institutions
- 2) Validates 800-character chunking with 100-character overlap as optimal for educational documents
- 3) Establishes all-MiniLM-L6-v2 as viable embedding model for production college chatbots
- 4) Provides comprehensive evaluation framework for educational chatbot systems
- 5) Enables practical deployment guidance for educational practitioners

Genie Assistant proves that sophisticated AI capabilities need not require expensive infrastructure. The open-source implementation maintains institutional autonomy while delivering capabilities exceeding commercial solution. Implementation across Indian colleges and universities can enhance experience, operational efficiency, and institutional information management.

## REFERENCES

- [1] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, 30, 5998-6008.
- [2] Devlin, J., Chang, M. w., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
- [3] Raffel, C., Shazeer, N., Roberts, A., et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-To-Text Transformer: *journal of Machine Learning Research*, 21(140),1-67.
- [4] Lewis, p., perez, e., Piktus, A., et al. (2020). Retrieval-Augmented Generation for knowledge-Intensive NLP Tasks. *Advances in Neural Information processing System*, 33, 9459-9474.
- [5] Singh, V., & Verma, R. (2023). Open-Source Language Models for Document-Based Question Answering in Indian Educational Contexts. *ACM Transactions on Asian and Low-Resource Language Information*
- [6] Kumar, A., Gupta, S., & Verma, p. (2023). Semantic Search Using Vector Embeddings for Institutional Knowledge Management. *Journal of Information Technology and Education*, 18(2), 123-145.
- [7] Thompson, J., Williams, R., & Davis, K. (2024), Privacy -Preserving Architectures in Educational Technology. *International Journal of Information Management*, 72, 102-119.
- [8] Johnson, m., & Lee, C. (2024). Adoption Of AI Chatbots in Educational Institutions: impact on Support Services and User Satisfaction. *Computers &Education*, 189, 104-125.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)