Evaluation Report: GIKI Prospectus Q&A Chatbot using Retrieval-Augmented Generation (RAG)

Iftikhar Ali

August 26, 2025

Contents

Chapter 1

Introduction

This evaluation report assesses the performance of the implemented **GIKI Prospectus Q&A Chatbot**, developed as per the project problem statement. The chatbot allows users to upload GIKI-related documents (e.g., UG Prospectus 2024) and query them using a Retrieval-Augmented Generation (RAG) pipeline for accurate, context-aware responses. The report compares the system's answers to original document content, evaluates accuracy, and provides an analysis of strengths and limitations.

Key project objectives evaluated

- Document ingestion and chunking.

- Efficient retrieval and generation using embeddings and LLM.

- User interface for interaction.

- Response accuracy and adherence to document content (no hallucinations).

The chatbot was tested with one example: querying the fee structure for the undergraduate Computer Science (CS) program from the *UG-Prospectus-2024.pdf* document.

Chapter 2

Methodology

2.1 Implementation Overview

The chatbot was built using Python with the following tools and technologies:

Document Ingestion

- Supports PDF, DOCX, TXT, PPTX, CSV, and XLSX.

- For PDFs, multiple extraction methods are used (PyPDF2, PyMuPDF, pdfplumber).

- Other formats handled via LangChain loaders or custom parsing.

Embedding & Vector Database

- Embedding Model: `sentence-transformers/all-MiniLM-L6-v2`.

- Text Splitting: SemanticChunker with fallback RecursiveCharacterTextSplitter (chunk size: 1000, overlap: 200).

- Vector Database: FAISS, storing chunks with metadata.

Retrieval-Augmented Generation (RAG)

- Retrieves top-8 relevant chunks (k=8).

- LLM: Llama3-8b-8192 via Groq API (temperature: 0.1, max tokens: 512).

- ConversationalRetrievalChain with custom prompt template to ensure document-based responses.

- Prompt emphasizes no hallucinations: *"If the answer is not found in the documents, reply: 'This information is not available in the uploaded documents.'"*

Chatbot Interface

- Framework: Gradio with custom UI.

- Features: Multiple document uploads, English/Urdu toggle (Urdu is placeholder).

- Missing: Source citation, export options.

- Styling: Modern CSS, avatars, clear button.

Deviations and Additions

- Supports more file types.

- Urdu translation is placeholder only.

- No automated evaluation framework (manual testing).

- Deployment: Local via Gradio launch.

2.2 Testing Approach

- **Documents Used:** UG-Prospectus-2024.pdf (282 pages).

- **Test Cases:** One provided (fee structure query).

Evaluation Criteria:

a) Accuracy (1–5 scale).

b) Completeness.

c) Hallucinations.

d) Formatting.

e) Manual usability feedback.

Chapter 3

Test Results

The table summarizes the test case comparing the chatbot's answer with the original source text.

Test Question	System Answer (Summary)	Original Source Text	Evaluation
What is the fee structure of CS undergraduate?	- Tuition: Annual foreign fee: US$ 405,000; Local: Rs. 467,500 (2026-27), Rs. 492,500 (2027-28). - Admin Charges: 5% of tuition (waived if lump sum). - Admission Fee: Rs. 65,000 / US$ 680. - Security Deposit: Rs. 40,000. - Convocation Fee: Rs. 17,500. - Mess Security: Rs. 15,000. - Refund Policy: Full before joining; partial within 7–15 days.	- Semester Fee (Tuition + Accommodation). - Tuition: Rs. 447,500 (2025-26), Rs. 467,500 (2026-27), Rs. 492,500 (2027-28). - Foreign Students: US$ 3,000. - Admin Charges: 5% of semester tuition. - Admission Fee: Rs. 65,000 / US$ 680. - Security Deposit: Rs. 40,000. - Convocation Fee: Rs. 17,500. - Mess Security: Rs. 15,000.	3/5 (Partially accurate).

Overall Metrics

- Total Tests: 1

- Average Accuracy: 3/5

- Hallucination Rate: Moderate

- Response Time: <10s

- Usability: Intuitive UI

Chapter 4

Analysis

Strengths

- Effective retrieval of relevant chunks.

- Robust ingestion pipeline.

- Professional response formatting.

- No complete hallucinations.

Limitations

- Hallucinations in foreign tuition fees and refund policies.

- No source citations in responses.

- Urdu toggle is placeholder.

- Only one test case evaluated.

Chapter 5

Conclusion

The chatbot meets core requirements for a RAG-based system, with effective document processing and retrieval. Accuracy was moderate (3/5 in tested case) due to hallucinations. The system is functional but requires refinements in prompt engineering, chunking, and evaluation for production readiness.