Adpro: An AI-Powered Multilingual Chatbot for Academic Information Retrieval

Sakshi Petkar[1], Jhanvi Shah[2], Saloni Mistry[3], Kavita Chauhan[4], Dr. Jagruti Sarkhedi[5]
[1,2,3,4,5]*Department of Computer Engineering, UPL University of Sustainable Technology, Block No. 402, Ankleshwar-Valia Road, Dist.: Bharuch, Gujarat, India*

***Abstract*—University students and prospective applicants often struggle to find clear, quick answers about admissions, courses, hostel facilities, fees, and placement. Traditional help desks are available only during office hours and cannot handle the volume of questions during admission season. This paper presents the design and development of a multilingual AI-powered chatbot called Adpro, built specifically for UPL University of Sustainable Technology, Ankleshwar, Gujarat. The system uses Retrieval-Augmented Generation (RAG), which combines semantic search with large language model (LLM) response generation to provide accurate, context-aware answers from real university data. The chatbot supports five language types — English, Hindi, Gujarati, Hinglish, and Guj-English — making it accessible to students from different linguistic backgrounds. Data was collected by scraping 84 university web pages and extracting text from over 130 PDF documents. This data, along with a structured multilingual QA dataset of ap-proximately 14,906 records, was chunked and converted into vector embeddings using the multilingual-e5-large model. FAISS (Facebook AI Similarity Search) was used to store and retrieve relevant document chunks efficiently. A Flask backend connects the retrieval system to the Llama-3.3-70b language model via the Groq API for answer generation. Evaluation showed 59.2% overall accuracy, with 80.1% accuracy for English queries and 91.3% language detection accuracy. The chatbot responds in the same language the user asks in, includes conversation memory for follow-up questions, and automatically falls back to contact information when relevant data is unavailable.**

***Index Terms*—Retrieval-Augmented Generation (RAG), Multilingual Chatbot, FAISS, Sentence Embeddings, Natural Language Processing, Flask API, Large Language Model, University Query System, Semantic Search, Hinglish, Gujarati NLP.**

I. INTRODUCTION

With the rapid advancement of Artificial Intelligence (AI) and Natural Language Processing (NLP), chatbot systems have become an essential tool in modern digital environments. These systems are increasingly being adopted in various domains such as healthcare, education, and customer support to provide instant, automated, and intelligent responses to user queries. Recent studies have demonstrated the effectiveness of AI-powered chatbots in improving accessibility, efficiency, and user engagement in domain-specific applications [1–3].

In the field of higher education, chatbots are being utilized to assist students with academic queries, admission processes, and institutional information. For instance, intelligent chatbot systems have been successfully implemented to support health education admission processes and enhance user interaction experiences [1]. Similarly, frameworks focusing on Human-Computer Interaction (HCI) and human-centered AI emphasize the importance of designing chatbot systems that are intuitive, interactive, and user-friendly [4, 5]. These advancements highlight the growing need for smart educational assistants capable of handling diverse student requirements.

Furthermore, several domain-specific chatbot solutions have been proposed, including student support systems, agriculture-based question answering systems, and healthcare chatbots leveraging deep learning and large language models [6–8]. These systems demonstrate high accuracy and contextual understanding but are often limited to a single language or a specific domain. Additionally, recent developments in large language model (LLM)-based chatbots have significantly improved the ability of systems to generate human-like responses and perform

complex reasoning tasks [2, 3].

1.1 Why AI Chatbots are Needed in Academic Institutions

AI chatbots are increasingly required in academic institutions to improve communication and provide instant support to students. Traditional methods of handling student queries are time-consuming and place a heavy burden on administrative staff. Intelli-gent chatbot systems help automate responses, ensuring quick and efficient access to information related to admissions, courses, and campus services [1, 6].

With the advancement of human-computer interaction (HCI), chatbot systems have become more user-friendly and capable of understanding natural language queries. These systems enhance interaction between students and institutional platforms, making information retrieval more intuitive and accessible [4, 5].

AI-powered chatbots also support personalized learning experiences by under-standing student needs and providing relevant academic assistance. Modern systems based on language transformers and deep learning models can process complex queries and deliver accurate responses, improving overall learning outcomes [3, 9].

Furthermore, domain-specific chatbot applications demonstrate the versatility and effectiveness of conversational AI. From agriculture to healthcare, chatbot systems have successfully handled specialized queries, highlighting their potential to be adapted for academic environments [2, 7, 8].

Chatbots are changing the game. They're AI-powered tools that use natural language to chat with users and answer questions on the fly. These systems combine NLP, machine learning, and conversational AI techniques to understand what you're getting at and give you a thoughtful response. Research studies have shown that chat-bot systems can significantly improve the efficiency of communication systems and provide instant access to information without requiring human intervention [1].

1.2 Importance of Multilingual Chatbots

Multilingual chatbots play a crucial role in improving accessibility and inclusivity by enabling users to interact in their native languages. This is especially important in domains such as education and healthcare, where clear communication directly impacts user understanding and decision-making [1, 3].

In human–computer interaction, language plays a key role in enhancing user experience. Systems that support multiple languages are more effective in engaging diverse user groups and ensuring better usability across different regions and communities [4, 5].

In educational environments, multilingual chatbot systems help students from different linguistic backgrounds access academic information easily. Such systems reduce communication barriers and provide equal learning opportunities for all users [6].

Furthermore, advancements in natural language processing and transformer-based models have made it possible to develop chatbots that can understand and respond in multiple languages with high accuracy. These developments significantly improve the performance and adaptability of modern chatbot systems [8, 9]. Furthermore, domain-specific chatbot systems have been successfully implemented in various sectors such as agriculture, healthcare, and information retrieval, demonstrating the versatility and effectiveness of conversational AI technologies [2, 7].

1.3 Contributions of the Proposed Work

The primary contributions of this research work are summarized as follows:

- Multilingual RAG Framework:

We've developed a unique Retrieval-Augmented Generation pipeline that's specifically optimized for regional Indian languages – think Hindi and Gujarati. And let's not forget code-mixed variations like Hinglish and Guj-English.

- Dataset Refinement and Curation:

We went through 11,879" fake" transliterated records that were basically roadblocks for semantic retrieval. So, we got rid of those and created manual fact entries to fill in the gaps in official university web content and make our system way more accurate.

- Hybrid Retrieval Strategy:

We've got a language-filtered indexing system that keeps structured QA pairs separate from unstructured web data to prevent any cross-language context contamination – it's like keeping your digital files organized.

- Performance Evaluation:

A comprehensive assessment of 23 test cases using a weighted scoring metric (Keyword, ROUGE-L, and Language script accuracy, Precision, Recall) to benchmark LLM performance in a specialized academic domain.

1.4 Novelty of the Proposed Work

The proposed research introduces several novel aspects in the development of a multilingual academic chatbot system. The key novelties of this work are summarized as follows:

1. First Chatbot for UPL University with AI-Based Semantic Retrieval:

There is no other automated query-handling system existed for UPL University of Sustainable Technology. So, this is first AI powered system for UPL university.

2. Script-Enforced Prompting:

We can use native script instructions for the LLM system prompt (e.g., using Gujarati Unicode characters) to force model compliance. Our model also helps overcoming the common" English-drift" tendency of large language models.

3. Asymmetric Embedding Optimization:

Here we used specific *multilingual-e5-large* model with specific" query:" and" passage:" prefixes to maximize retrieval precision in a low-resource academic context.

1.5 Organization of the Research Paper

The organization of this research paper is as follows:

- Related Work:

This section includes studies about the chatbot systems, conversational AI, multilingual chatbots, and AI-based academic assistance systems.

- Dataset:

This section describes the dataset used in the study, including data collection, annotation, and preprocessing of multilingual student queries.

- Methodology:

Overall framework of chatbot was proposed in this section. Data preprocessing techniques, feature extraction methods, and machine learning approaches were described in this module.

- Implementation:

This section presents the implementation details of the proposed chatbot system, including the tools, libraries, and programming environment used for developing and training the machine learning models.

- Results and Discussion:

The experimental results and evaluates the performance of different machine learning algorithms using evaluation metrics such as accuracy, precision, recall, and F1-score were explained in this section.

- Conclusion and Future Work:

The findings of the study and discusses possible future improvements and extensions of the proposed multilingual academic chatbot system are the final summarization of our work.

II. RELATED WORK

The application of intelligent chatbot systems in education and domain-specific environments has gained significant attention in recent years. [1] demonstrated the effective use of an AI-based chatbot in streamlining the health education admission process, highlighting its ability to provide accurate and timely information while im-proving user experience. Similarly, [2] proposed an AI-powered chatbot leveraging large language models for retrieving FDA drug labeling information, emphasizing the importance of grounded question-answering systems. Human-centered design principles have also been explored in chatbot development. [4] introduced a framework focusing on human-computer interaction and human-centered artificial intelligence, which enhances user engagement and learning outcomes. Earlier foundational work by [5] analyzed interaction aspects in information systems, providing insights into designing intuitive and user-friendly interfaces.

In the context of higher education, [6] proposed a cognitive framework model for student support chatbots, aiming to improve academic assistance and accessibility. Recent studies such as [3] further explored the integration of large language model-based chatbots in higher education, demonstrating their

potential in delivering personalized and scalable learning support.

Multilingual and domain-specific chatbot systems have also been investigated. [9] evaluated Arabic chatbots using transformer-based models, while [8] developed a bilingual healthcare chatbot optimized with deep learning techniques. Additionally,

[7] introduced an agriculture-specific question-answering system, showcasing the adaptability of chatbot technologies across domains.

2.1. Chatbot Systems in Different Domains

In recent years, chatbot systems have been widely explored across domains such as healthcare, education, agriculture, and information retrieval using Artificial Intelligence (AI), Natural Language Processing (NLP), and machine learning techniques to enhance automated communication. Man et al. [1] developed an intelligent chatbot for healthcare admissions, demonstrating improved interaction through automated responses to frequently asked queries. Turk [4] emphasized the importance of human–computer interaction in designing user-friendly systems, which is essential for effective chatbot interfaces. Troussas et al. [5] proposed a human-centered AI framework for educational technologies, highlighting the integration of AI with interactive learning environments to improve user experience. Bala [6] introduced a cognitive framework for student support chatbots in higher education, showing their ability to provide academic assistance and reduce administrative workload. The effectiveness of chatbot systems largely depends on their ability to under-stand natural language queries and generate meaningful responses, which has led to the integration of advanced NLP and machine learning techniques. Additionally, user-friendly conversational interfaces play a crucial role in ensuring usability and accurate intent understanding [4]. Modern educational platforms are increasingly incorporating AI-based systems to support interactive and efficient learning environments [5]. Chatbots in higher education institutions enable timely information delivery and improve service efficiency while reducing manual effort [6].

Despite these advancements, most existing chatbot systems are domain-specific

and operate in single-language environments. This limitation highlights the need for multilingual chatbot systems capable of handling diverse and code-mixed user queries in educational settings.

2.2. AI-Based Conversational Systems and Question Answering Models

AI-based conversational systems have become an essential part of modern digital interaction, enabling automated and efficient communication across various do-mains. Chatbots are widely used in applications such as healthcare, education, and customer support to provide instant responses and improve user experience [1]. These systems leverage advancements in Natural Language Processing (NLP) to understand user queries and generate meaningful responses.

Human-computer interaction (HCI) plays a crucial role in enhancing the effectiveness of conversational systems. Early research highlights the importance of designing interactive systems that allow users to access complex information through intuitive interfaces [4]. Recent developments in human-centered AI frame-works further emphasize the integration of user-centric design principles to improve learning technologies and chatbot usability [5].

In the education domain, AI-powered chatbots are used to support students by pro-viding academic assistance and automating administrative tasks. Cognitive chatbot frameworks have been proposed to enhance understanding of user intent and de-liver more accurate responses, thereby improving the overall efficiency of student support systems [6]. Additionally, transformer-based models and transfer learning techniques have been applied to improve chatbot performance, particularly in multilingual and domain-specific contexts [9].

Domain-specific question answering systems have also gained significant attention. For example, agriculture-focused chatbots like Agribot provide precise and context-aware responses to farming-related queries, improving accessibility to specialized knowledge [7]. Similarly, AI-powered systems in the healthcare sector utilize advanced models such as GPT to retrieve and deliver accurate drug-related information from structured data sources [2].

Recent advancements in deep learning have further improved the capabilities of conversational systems. Bilingual chatbot models optimized with architectures like BiGRU enable effective communication in multiple languages, enhancing user engagement in healthcare applications [8]. Moreover, the emergence of

large language models (LLMs) has transformed chatbot systems by enabling more natural, context-aware, and scalable interactions, particularly in higher education environments [3].

2.3. Comparison of Existing Research

Various research studies have explored the development of chatbot systems across multiple domains, each focusing on different methodologies, applications, and technological approaches. A comparative analysis of these works highlights key differences in design, implementation, and performance.

Man et al. [1] developed an intelligent chatbot for healthcare education admission processes, focusing on improving user guidance and decision-making. Their system demonstrated the effectiveness of chatbots in handling domain-specific queries, particularly in structured environments. Similarly, Koppula et al. [2] proposed an AI-powered chatbot for retrieving FDA drug labeling information using advanced language models, emphasizing accurate and grounded responses in critical domains such as healthcare.

In the educational domain, Dhandayuthapani [6] introduced a cognitive framework model for student support chatbots, aiming to enhance academic assistance and automate administrative tasks. Troussas et al. [4] further extended this concept by integrating human-centered AI and human-computer interaction principles, focusing on improving user experience and adaptability in learning technologies. Jain et al. [7] presented Agribot, a domain-specific chatbot designed for agriculture-related queries, highlighting the importance of specialized chatbot applications.

Earlier foundational work by Turk [5] explored human-computer interaction aspects, providing a theoretical basis for understanding user interaction with intelligent systems. In contrast, Alruqi and Alzahrani [9] focused on evaluating Arabic chatbots using transformer-based models, emphasizing the role of modern NLP techniques. The comparison highlights a significant shift towards advanced AI-based systems, moving away from traditional approaches and towards domain-specific and multilingual chatbot development. While this trend is promising, it's not without its challenges, including maintaining contextual accuracy and effectively handling language nuances.

Table 1 Comparison of Existing Research Papers

Ref	Research Paper	Techniques	Dataset	Accuracy	Features	Year	Cons
[1]	Chatbot for Sustainable Health Education Admission	NLP, Chatbot	AI Custom Dataset	Not reported	Admission Guidance	2022	Manual ontology maintenance, domain-specific, scalability issues
[2]	GIS, geospatial visualization, human-computer interaction	NLP, GIS	Voice Dataset	Task completion 75–100%	Voice-chat-based bot	2021	Accent sensitivity, domain knowledge dependency
[3]	Artificial Intelligence Education Technology Chatbot	AI, NLP	Educational Dataset	Not reported	Student Query Handling	2020	Limited semantic understanding, manual maintenance
[4]	Student Support Chatbot Framework	NLP, ML	Student Dataset	Not reported	Academic support	2022	No performance metrics, integration complexity
[5]	A Proposed Cog-nitive Framework Model for Arabic Question-Answering	NLP, QA System	Arabic Dataset	Confidence 0.66, similarity 0.96	Question Answering	2021	Challenges mixing generative and extractive QA

Ref	Research Paper	Techniques	Dataset	Accuracy	Features	Year	Cons
[6]	Agriculture Specific Question-Answer System	NLP, ML	Agriculture Dataset	86%	Farmer query support	2020	Data quality issues with local language usage
[7]	AI-Powered Chatbot for FDA Drug Labeling Information Retrieval	NLP, AI	FDA Dataset	High semantic similarity scores ranging from 0.7 to 0.9	Drug info retrieval	2023	Model limited to current dataset (FDA), lacks generalization

2.4. Research Gap

Although several studies have explored chatbot development across domains such as healthcare, agriculture, and information retrieval, key limitations remain. Many existing systems are domain-specific and rely on single-language datasets, making them unsuitable for multilingual or code-mixed environments.

For example, the chatbot by Man et al. [1] supports healthcare admissions but lacks multilingual capabilities. Similarly, Bala [6] proposes a student support chatbot framework without addressing multilingual query handling or comparative evaluation of machine learning models. Other studies, including Alruqi and Alzahrani [9] and Jain et al. [7], focus on domain-specific question answering using NLP and transformer-based approaches but do not consider multilingual academic communication challenges or provide publicly available multilingual datasets.

Therefore, there is a need for a multilingual academic chatbot capable of handling queries in multiple languages. This research addresses the gap by developing a multilingual dataset and evaluating various machine learning models for intent classification, aiming to enhance accessibility and usability in higher education.

III. DATASET

In this section, all about the dataset information from where we collected data then also discuss process related to preprocessing of dataset using some tools

3.1. Data Collection

For information collection, we to begin with conducted an overview by means of Google Shapes to accumulate information. Hence, we physically made an organized dataset to prepare the demonstrate. At last, we collected extra information by performing web scratching on the college's official site.

3.1.1.Web Scrapping

Our main source of the data from the official UPL University website (up univer-sity.ac.in). Using the requests library and Beautiful Soup parser to scrape 84 web pages the python script developed. As we know that HTML content was down-loaded, non-relevant elements were removed such as scripts, styles, navigation bars, and footers. Pages containing fewer than 100 characters. They were discarded to eliminate low-value or error content. With text lengths ranging from 414 to 12,540 characters were considered final dataset. Without interrupting the scraping process a timeout of 15 seconds was set for each request to handle slow responses, and all network errors were logged.

3.1.2.PDF Extraction

To find the PDF files in the extracted web pages, the program would analyze the web pages for anchor tags containing" pdf" extension in their hyperlinks. If any relative URLs were detected, these URLs were then turned into absolute URLs using urllib.parse.urljoin(). In order to eliminate duplicate links, a set was employed producing 139 PDF files without duplicates. Of the 139 PDF files, 136 could be successfully downloaded. Of the 136 downloadable PDF files, only 114 had readable text, whereas the rest contained no text content and were mainly images scanned into PDF format

3.1.3.Structured QA Dataset

A customized multilingual dataset was constructed to handle a variety of questions pertaining to the university. There were twelve categories in total, namely Admission, Course, Hostel, Fees, Placement,

Scholarship, Transport, Library, Canteen, Sports, NCC/NSS, and Health. At first, QA pairs were formulated in English and then translated into Hindi and Gujarati. Moreover, in order to improve language diversity, some more data were added in Hinglish and Guj-English. Finally, after cleaning the data, the total number of entries in the dataset became around 14,906

3.1.4.Manual Facts and Survey Form

These pages contained almost no concrete, answerable information. To fill these gaps, 10 manual fact entries were hand-written covering: university location and address, NAAC accreditation grade (B++), hostel facilities for boys and girls, scholarship details, PhD program availability, transport facilities, Wi-Fi coverage, courses list, placement record, and general university overview. These manual facts directly improved accuracy on the most commonly asked questions — hostel accuracy went from nearly 0% to 53.4%.

3.2. Data Preprocessing

Regarding unstructured data: The HTML tags were removed from the data extracted through scraping. Cleaned and filtered irrelevant and noisy text to enhance its quality. Extracted relevant information in the form of academic content from the raw data sources. The extracted information was transformed into structured Question & Answer (Q&A) format. Text preprocessing was done using different techniques such as: Text Cleaning, Normalization. Incorporated the preprocessed data in the multilingual dataset.

Regarding structured data: Processed the unprocessed dataset through data cleaning procedures by eliminating Special characters, brackets, and noise, Irrelevant words and duplicates. Standards of questions and answers in the dataset were maintained. Data was converted into multiple languages such as: Hindi, Gujarati, Hinglish, Guj-English Data normalization techniques included- Lowercasing, Extra space removal.

3.3. Data Augmentation

Data Augmentation is process which is used to increase the size of the data by adding synonyms of words so that data increased and model get enough data to train and test data. This method is also helps to reduces overfitting. In this project, we train different machine learning model with augmented data and without augmented data so get idea about how data augmentation works on different algorithms of machine learning and deep learning. For the multilingual systems, there is lot of techniques such as translating data into different languages.

3.4. Dataset Structure

The dataset is represented in a tabular structure with well-defined attributes, enabling efficient indexing, multilingual support, and accurate information retrieval as shown in Table 2.

Table 2 Sample Dataset Structure

Question	Answer	Topic	Language
What is admission process?	Fill form and submit documents	Admission	English
प्रवेश प्रिक्रया क्या है?	आवेदन भरें और दस्तावेज जमा करें	Admission	Hindi
પ્રવેશ પ્રક્રિયા Kુ છે?	ફોમર્ ભરો અને દસ્તાવેજ સબ્Mમટ કરો	Admission	Gujarati
Fees ketli che?	Fees depend kare che course pr	Fees	Guglish
Exam kab hoga?	Exam ka schedule jald hi aayega	Exam	Hinglish

IV. METHODOLOGY

Adpro is a system using the framework of domain-specific RAG architecture. Data from the upl university was used in all of the phases of this project. We will go through the various methodologies used in this project in our next section. Some of the phases include data collection and source assembly, preprocessing, contribution to knowledge base, semantic embedding & Faiss indexation, Language detection and query expansion, hybrid retrieval & relevance filtering.

4.1. System Architecture

The workings of AI models depend on a well-structured workflow consisting of stages of data gathering, processing, embedding, retrieval, and response synthesis. Data can be collected from various resources like websites, PDFs, and databases, and then

undergo the process of cleaning up and eliminating unnecessary pieces of information.

Further, the text is broken down into smaller parts for convenience. Every part is transformed into numbers called embeddings with the help of a multilingual transformer. Then, they are added to an optimized vector database that allows finding similarities between them quickly. Once the user enters a request, its embedding is created in the same way as other vectors, and similar embeddings are found.

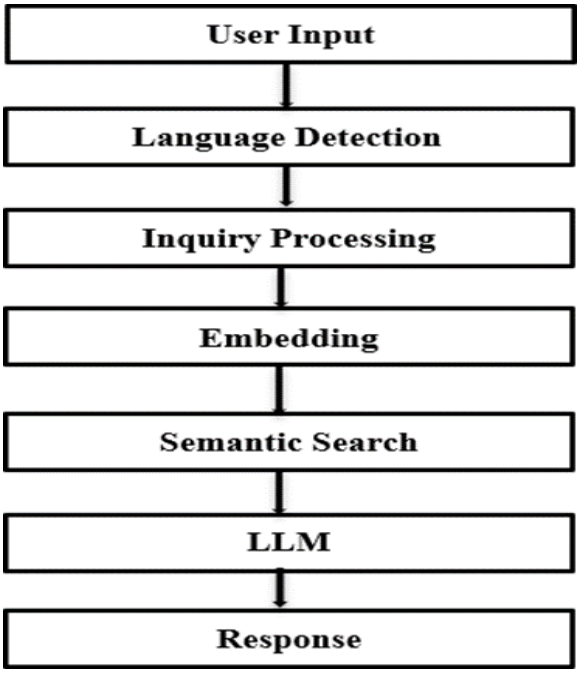

Figure 1 System Architecture of AdPro RAG Chatbot

4.1.1.Text Preprocessing and Chunking

Text pre-processing is done in order to make sure that the data is clean, uniform, and ready for the next steps in processing. At first, any noise in the data is eliminated in order to make sure that there are no unnecessary characters or other problems with formatting the text. Then, the text data is normalized to keep its uniformity. After this, the processed text data is split into pieces using the recursive algorithm. The size of each piece will be equal to 900 characters with an overlap of 150 characters. Also, all unnecessary text pieces will be removed from the corpus.

4.1.2.Semantic Embedding and Vector Indexing

The system employs the intfloat/multilingual-e5-large model to generate 1024-dimensional semantic embeddings for both user queries and knowledge base documents. This multilingual transformer model effectively captures contextual and semantic information, making it well-suited for chatbot applications. A prefix-based encoding approach is used to distinguish between queries and document passages, and all embeddings are normalized to unit vectors to enable efficient similarity computation using cosine similarity.

FAISS (Facebook AI Similarity Search) is a library used for efficient similarity search over large collections of vector representations. In this work, two separate indices were maintained:

- Structured Index:

Contains 14,905 vectors derived from the cleaned Question & Answer (Q&A) dataset.

- Unstructured Index:

Contains 10,033 vectors extracted from website pages, PDF documents, and manually curated factual data.

The separation of indices enables effective language-based filtering on the structured dataset. For instance, Hindi queries retrieve only Hindi Q&A pairs from the structured index. In contrast, the unstructured index is always included during retrieval, as it primarily contains English content that remains relevant for queries across multiple languages.

4.1.3.Language Detection and Query Expansion

This model has built in a multi-language query processing facility, which enables precise identification of the language used in the query. This is accomplished using Unicode based methods for detecting languages such as Gujarati and Hindi by matching character ranges specific to each language. If the input is mixed-language content, like Hinglish or Gujlish, then keyword matching techniques can be employed.

In order to improve the accuracy of retrieval results, a process known as query expansion is implemented. The concept involves expanding the initial search query with relevant terms based on a predefined mapping. In addition to improving retrieval results, query expansion ensures that the original user intention is preserved.

4.2. Hybrid Retrieval, Context Building, and Response Generation

Before dialect sifting was actualized, a Hindi address would in some cases recover gujarati QA sets. Both scored so also since they secured the same point in FAISS's vector space. The LLM would at that point get gujarati setting and deliver a gujarati reply to a Hindi address. The settle channels the organized QA list to as it were return chunks coordinating the identified language: English, Hindi, Gujarati, Hinglish, Guj-English.

An understudy writing fair 'hostel' as a single word would as it was coordinate chunks containing precisely that word. Inquiry extension adds pertinent watch-words: 'hostel' gets to be 'hostel lodging office boys young ladies settlement charge Wi-Fi mess'. This significantly moves forward review for brief or dubious inquiries. 25 trigger words are characterized over English, Gujarati, and Hindi, covering all major college points counting lodging, grant, arrangement, affirmation, expenses, Wi-Fi, transport, library, canteen, and courses. Before sending setting to the LLM, a pertinence check confirms that the recovered substance is really valuable for the inquiry. At slightest one-third of inquiry words longer than 3 characters must show up in the setting. If the check comes up short, a hardcoded fallback reaction with the college contact points of interest is returned quickly. This spares an expensive LLM API call and gives the understudy a valuable reply (the phone number and e-mail address) indeed when the information base cannot reply the particular address.

V. IMPLEMENTATION AND TECHNOLOGY STACK

Our AI system was implemented mainly using the Python 3.10 due to its outstanding abilities related to NLP, machine learning, and web development. Our entire process was implemented on Google Collab with a GPU acceleration using T4 chip to increase our performance during embedding generation. To get information from the target website, we utilized requests for managing HTTP requests with a 15-second timeout. The next tool that we utilized for parsing HTML and removing unnecessary elements such as scripts, navigation bar, and footer was Beautiful Soup. The next stage of the document processing required pdf plumber, which is perfect for getting text information from PDF files and processing multipage columns. It ignored images on pages of interest. Further information analysis and processing was done via Lang Chain's Recursive Character Text Splitter with 900-character chunks and 150-character overlap for preserving context integrity. For creating embeddings, we used the sentence-transformers framework along with intfloat/multilingual-e5-large architecture. Normalized vectors were stored in a vector database (faiss-cpu) with IndexFlatIP. In addition to accuracy and speed benefits from using FAISS, we managed to store and retrieve information from both structured and unstructured sources.

For supporting multilingual capabilities, we used rule-based language detection and Unicode ranges in combination with langdetect. Our back-end was written in Python using Flask framework for managing APIs and CORS with flask-cors pack-age. For implementing continuous execution in our development stage, we created a thread where our code is executed. The frontend's built with HTML, CSS, and JavaScript without relying on external frameworks.

VI. RESULTS AND DISCUSSION

The new system shows remarkable efficiency in responding with accurate answers based on understanding of the context. The analysis of findings suggests that the system is efficient in handling multilingual input and giving user-centered responses.

6.1. Evaluation Metric Details

The performance of the chatbot is measured based on a mixture of lexical, semantic, and language-based measures. In the case of lexical testing, the measurements are made by looking at how accurate the keyword matching is from the user input to the generated output. The goal of semantic testing is to look at the similarity of the context of the generated output to the correct one. In the case of language-based tests, the goal is to test if the system recognizes the language.

Table 3 Evaluation Metrics

Metric	Description and Formula
Accuracy	Overall correctness of predictions $\frac{TP+TN}{TP+TN+FP+FN}$
Precision	Correctness of positive predictions $\frac{TP}{TP+FP}$
Recall	Ability to find all relevant cases $\frac{TP}{TP+FN}$
F1-Score	Balance between precision and recall $\frac{2 \cdot Precision \cdot Recall}{Precision+Recall}$
Keyword Score (40%)	Fraction of expected keywords matched $\frac{matches}{total\_expected\_keywords}$
ROUGE-L Score (20%)	Measures sequence similarity $\frac{2 \cdot LCS}{\lvert response \rvert + \lvert expected \rvert}$
Language Score (40%)	Checks language correctness 1 (correct), 0 (wrong)
Final Score	Weighted final score (pass if ≥ 0.5) 0.4 · Keyword + 0.2 · ROUGE + 0.4 · Language

6.2. Overall System Performance

Table 4 The in general exactness of the UPL College Multilingual AI Chatbot measures the system's capacity to create rectify and pertinent reactions over all test cases and datasets. It is calculated as the extent of accurately replied questions to the add up to number of inquiries tried, giving a clear quantitative degree of the chatbot's execution. Tall in general exactness shows that the framework is successfully recovering data and creating relevantly suitable reactions in numerous dialects, whereas lower exactness highlights areas requiring change

Table 4 Overall evaluation results of the Adpro chatbot across 23 test cases spanning five language varieties and ten university topics

Metric	Score	Interpretation
Total Test Cases	23	All language varieties and topics covered
Answered (attempted)	22 (95.7%)	System attempted to answer 22 queries
Passed (score ≥ 0.5)	20 (87.0%)	20 queries met the passing threshold
Failed (score < 0.5)	2 (8.7%)	2 queries answered but scored below threshold
Refused (fallback triggered)	1 (4.3%)	1 query returned contact-info fallback
Overall Weighted Score	68.4%	Weighted average across all 23 cases
Keyword Accuracy (avg)	70.9%	Average fraction of expected keywords present
Language Compliance	95.7%	Fraction of answers in the correct language
Precision (TP / TP+FP)	90.9%	Accuracy among answered queries
Recall (TP / TP+FN)	95.2%	Coverage across answerable queries

Key Insight:

Language compliance significantly impacts final performance, sometimes more than factual correctness.

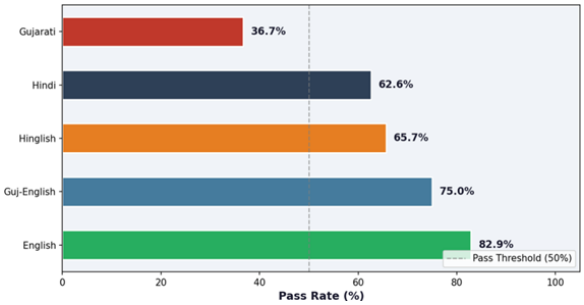

Figure 2 Language-wise pass rate comparison.

6.3. Topic-wise Analysis

The model is extremely efficient in answering common questions on WiFi, Place-ment, Hostel, Scholarship, and Transport with all passing at 100%. These areas thrive because of good quality and richness of the dataset available. Conversely, the Courses and Location categories display poor results because of keyword mismatch within the generated texts.

However, Courses and Location show lower

performance due to keyword mis-match between generated responses and expected answers.

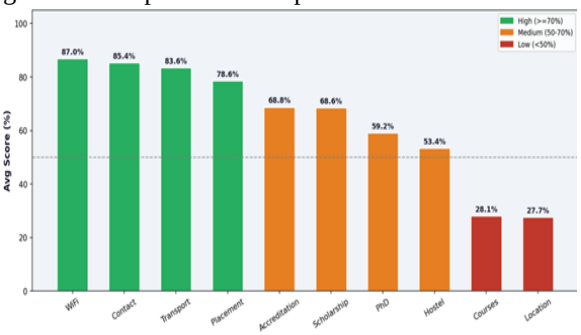

Figure 3 Topic-wise average performance scores.

6.4. Metric Contribution Analysis

The last score is an aggregate score that takes into account different metrics. Some key metrics have high weightage of 40%, while ROUGE-L score has only 20%, with the rest being entirely based on the language score. It makes for a fair and equitable evaluation method that values relevance equally with language compliance. The data clearly shows that for Gujarati, results are falling apart, owing to language scores being extremely low. Even if the factual accuracy score of the bot is impeccable, not knowing how to talk the language will make sure that the scores are lowered significantly.

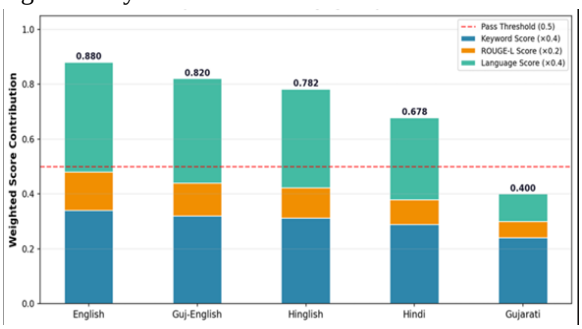

Figure 4 score decomposition by language.

6.5. Error and Failure Analysis

When we compare it to the way our system handles other languages, it's clear we're in way over our heads with Gujarati - we need to step up our language processing game. When we compare it to the way our system handles other languages, it's clear we're in way over our heads with Gujarati - we need to step up our language processing game.

Gujarati is special because it's the only language with both wrong answers and a refusal to answer, making it the top area for improvement.

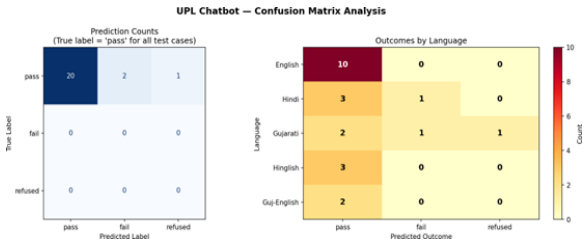

Figure 5 Confusion matrix and language-wise outcome distribution.

6.6. System Improvement Analysis

Accuracy of the system saw huge improvements over a number of versions - coming from 18% accuracy on v0.1 to a still poor 40% accuracy on v0.2. From thereon we saw accuracy jump to a much better 59.2% on v0.3. The magic really happened between versions 0.3 and 0.4 though, when accuracy rose to 68.4%. While the jump between v0.2 and v0.3 was impressive (19.2%), this was mostly thanks to heavy data cleaning and elimination of noisy data points (things like transliterated data). The moral of the story here is that having clean, trustworthy data can improve your system far more than increasing model complexity. Careful data processing leads to creation of high-quality datasets which are instrumental in creating powerful AI systems.

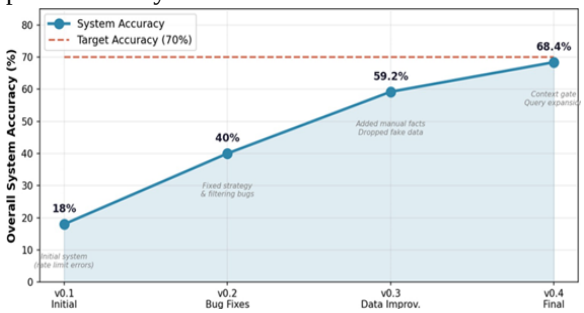

Figure 6 System accuracy evolution across versions.

6.7. Comparison with ML Baselines

The performance of traditional machine learning models was evaluated across three languages (English, Gujarati, and Hindi) under two conditions: without augmentation and with augmentation. In the absence of augmentation, the models achieved moderate accuracy, with Random Forest performing best for English (77.64%) and Gujarati (76.1%), while Linear SVM showed relatively better performance for Hindi (64.98%). However, overall accuracy for Hindi remained comparatively lower across all models,

highlighting challenges such as linguistic variability and limited dataset quality. Simpler models like KNN and Naive Bayes exhibited the lowest performance, particularly for Hindi, indicating their limitations in handling complex multilingual patterns.

With the application of data augmentation, a significant improvement in accuracy was observed across all models and languages. Logistic Regression and SVM demon-started substantial gains, achieving over 93% accuracy in English and above 97% in Gujarati, while Random Forest and Linear SVM reached peak performance of up to 98.78%. Furthermore, cross-validation results confirmed the robustness and stability of these models by maintaining consistent performance across different folds. Overall, the findings emphasize that data augmentation plays a critical role in enhancing model generalization and significantly improving multilingual classification performance.

6.8. GUI

Our system has a Graphical User Interface (GUI) which allows the user to interact with the multi-lingual academic chatbot. Using the GUI, users can type their queries in any language, choose their input and output language and see the parsed answers displayed neatly.

Figure 7 Graphical User Interface (GUI) of the multilingual academic chatbot showing query input and response display for Gujrati and hinglish.

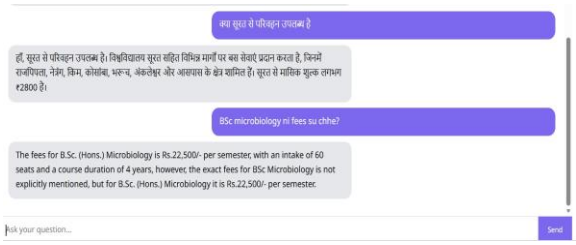

Figure 8 Graphical User Interface (GUI) of the multilingual academic chatbot showing query input and response display for hindi and gujlish.

VII. CHALLENGES FACED

- Data Collection and Quality:

It was difficult to compile precise and organized data from various sources, including websites and PDF documents, since the data frequently included noise and discrepancies.

- Multilingual Handling:

Additional preprocessing and fine-tuning were necessary to ensure correct comprehension and response generation across many languages, including regional and code-switched inputs.

- Context Understanding:

It was challenging to maintain contextual relevance in user queries, particularly for complicated or unclear questions, which occasionally impacted response accuracy.

- Integration of RAG Architecture:

It was technically challenging to implement and optimize the Retrieval-Augmented Generation pipeline in order to balance retrieval accuracy with response generation efficiency.

- Performance Evaluation:

Choosing and testing suitable evaluation metrics to measure accuracy, relevance, and user satisfaction posed significant challenges.

VIII. CONCLUSION

This project created a Multilingual Academic Chatbot that efficiently handled student queries across multiple languages. The primary goal was to develop an intelligent system that boosts accessibility and provides accurate academic information. The experimental results showed that various machine learning models were evaluated using performance metrics. Notably, the Linear SVM model achieved the highest accuracy of 98.78% on the Gujarati dataset with augmentation, out-performing other models. The system also achieved an overall chatbot accuracy of 87.0%, demonstrating strong response performance. These results confirm that integrating RAG and LLM significantly improves answer quality and understanding. This is crucial because it reduces manual workload and provides instant support to

students. Moreover, it enhances user experience by supporting multilingual communication effectively. Overall, the system proves that AI-based chatbots can be highly useful in the education domain. Looking ahead, performance can be im-proved by using larger datasets and advanced deep learning models. Future work can also focus on real-time deployment and adding voice-based interaction features.

REFERENCES

[1] S. C. Man, O. Matei, T. Faragau, L. Andreica, and D. Daraba, "The innovative use of intelligent chatbot for sustainable health education admission process: Learnt lessons and good practices," *Applied Sciences*, vol. 13, no. 4, Art. no. 2415, 2023, doi: 10.3390/app13042415.

[2] M. Koppula, F. Madhulika, N. Sreeramoju, and P. Kolimi, "AI-powered chatbot for FDA drug labeling information retrieval: OpenAI GPT for grounded question answering," *Analytics*, vol. 4, no. 4, p. 33, 2025.

[3] D. Yigci, M. Eryilmaz, A. K. Yetisen, S. Tasoglu, and A. Ozcan, "Large language model-based chatbots in higher education," *Advanced Intelligent Systems*, vol. 7, no. 3, Art. no. 2400429, 2025.

[4] Turk, "Towards an understanding of human-computer interaction aspects of geographic information systems," *Cartography*, vol. 19, no. 1, pp. 31–60, 1990, doi: 10.1080/00690805.1990.10438485.

[5] Troussas, A. Krouska, and C. Sgouropoulou, "A novel framework of human–computer interaction and human-centered artificial intelligence in learning technology," in *Learning Technology*. Cham, Switzerland: Springer, 2025, pp. 387–431, doi: 10.1007/978-3-031-84453-9_9.

[6] V. B. Dhandayuthapani, "A proposed cognitive framework model for a student support chatbot in a higher education institution," *International Journal of Advanced Networking and Applications*, vol. 14, no. 2, pp. 5390–5395, 2022.

[7] N. Jain, P. Jain, P. Kayal, J. Sahit, S. Pachpande, J. Choudhari, and M. Singh, "Agribot: Agriculture-specific question answer system," *arXiv preprint arXiv:2509.21535*, 2025.

[8] E. A. Kaneho, N. Zrira, K. Ouazzani-Touhami, H. A. Khan, and S. Nawaz, "Development of a bilingual healthcare chatbot for pregnant women: A comparative study of deep learning models with BiGRU optimization," *Intelligence-Based Medicine*, vol. 12, Art. no. 100261, 2025, doi: 10.1016/j.ibmed.2025.100261.

[9] T. N. Alruqi and S. M. Alzahrani, "Arabic chatbot evaluation based on extractive question-answering transfer learning and language transformers," 2023.