CUSTOMIZED NLP-BASED CHATBOT FOR EGS

DR. A.EMMANUEL PEO MARIADAS[1], A.MUKILARASI[2], R.SANDHIYA[3] , S.USHA[4]

[1]Project Supervisor, Assistant Professor,

[2,3,4] Final Year Student

[1,2,3,4] Department of Computer Science and Engineering,

[1,2,3,4] E.G.S. PILLAY ENGINEERING COLLEGE, Nagapattinam, India

Abstract - In recent years, conversational agents have gained significant attention in educational institutions for providing instant and automated information services. This paper presents the design and implementation of a customized college information chatbot developed using Natural Language Processing (NLP) and Machine Learning techniques. The proposed system aims to assist students, faculty, and visitors by answering college-related queries such as admissions, courses, departments, examinations, and placements in an efficient and user-friendly manner. The chatbot utilizes the NLTK library for text preprocessing tasks including tokenization, stopword removal, and stemming, which helps in improving the quality of input data. Feature extraction is performed using the TF-IDF vectorization technique, and intent classification is achieved through machine learning models implemented using scikit-learn. Based on the predicted intent, appropriate responses are retrieved from a predefined knowledge base. The system is lightweight, operates without internet dependency, and can be deployed on low-resource environments, making it suitable for academic institutions. Experimental results demonstrate that the chatbot provides accurate and relevant responses for domain-specific queries. The proposed approach offers a cost-effective and scalable solution for automating information access in college environments and serves as a foundation for future enhancements using advanced deep learning models.

Index Terms - Natural Language Processing, Chatbot, Machine Learning, Scikit-learn, NLTK, Institutional Assistant, Text Processing, AI Chatbot.

I. INTRODUCTION

The rapid advancement of artificial intelligence and natural language processing has significantly influenced the development of intelligent conversational systems. Chatbots have become an effective medium for providing instant and automated responses to user queries in various domains, including education, healthcare, banking, and customer support. In educational institutions, students and visitors frequently seek information related to admissions, courses, departments, examinations, academic schedules, and placement activities. Handling such repetitive queries manually is time-consuming and increases the workload of administrative staff.

Traditional methods of information dissemination, such as notice boards, websites, and help desks, often fail to provide real-time interaction and personalized responses. As a result, there is a growing need for intelligent systems that can understand natural language queries and deliver relevant information efficiently. Natural Language Processing (NLP) enables machines to analyze, understand, and interpret human language, thereby allowing users to interact with systems in a more natural and intuitive manner.

Recent advancements in machine learning have improved the performance of NLP-based systems by enabling automatic intent detection and text classification. However, large-scale transformer-based models require high computational resources and continuous internet connectivity, which may not be feasible for small or medium-level educational institutions. Hence, lightweight and efficient solutions using traditional NLP and machine learning techniques are more suitable for academic environments.

This project presents a customized college information chatbot developed using NLTK and scikit-learn. The proposed system focuses on understanding user intent through text preprocessing and machine learning-based classification techniques. It provides accurate responses from a predefined knowledge base and operates efficiently in a local environment. The system aims to improve accessibility to college-related information while reducing manual effort and operational cost.

II. PROPOSED SYSTEM

The proposed system introduces a customized chatbot designed specifically for institutional query handling. The system processes user queries, analyzes intent, and generates appropriate responses using trained models.

2.1 Objectives of the Proposed System

• To develop an automated query response system for college information.
• To implement NLP-based text understanding.
• To classify user queries using machine learning techniques.
• To provide fast and accurate responses using institutional datasets.
• To reduce manual workload in academic support services.

2.2 Core Modules of the Proposed System

• User Query Input Module
• Text Preprocessing Module
• Machine Learning Classification Module
• Response Generation Module
• Database Management Module

2.3 Working Principle

Step 1: User enters query through the interface.
Step 2: Text preprocessing is performed using NLTK.
Step 3: Features are extracted using NLP techniques.
Step 4: Query classification is performed using Scikit-learn.
Step 5: Relevant response is retrieved from the dataset.
Step 6: Response is displayed to the user.

2.4 Advantages of the Proposed System

• Provides instant response to student queries.
• Reduces manual administrative work.
• Domain-specific knowledge handling.
• Easy integration with institutional systems.
• Scalable and efficient solution.

III. SYSTEM ARCHITECTURE

The system follows a modular architecture designed to ensure scalability, security, and efficient response generation. The architecture integrates a chatbot interface, NLP-based query processing, a customized GPT model, and an institutional database.

The user interacts with the system through a web or mobile-based chatbot interface. Queries are authenticated and processed to identify intent and relevant keywords. The customized GPT module generates responses using both learned knowledge and real-time data retrieved from the EGS database.

System Architecture Diagram: Customized Generative Pre-Trained Transformer for EGS
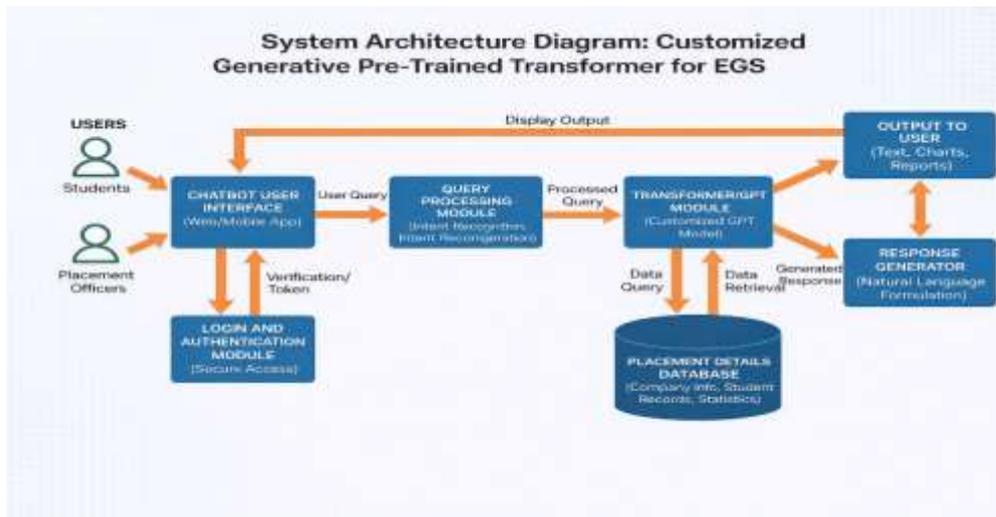
3.1 User Interface Layer

Provides interaction through:
• Student and staff login
• Query submission
• Response display

3.2 Query Processing Layer

Responsible for:
• Tokenization and preprocessing
• Intent recognition
• Context extraction

3.3 Transformer/GPT Layer

• Customized GPT model
• Fine-tuned using EGS data
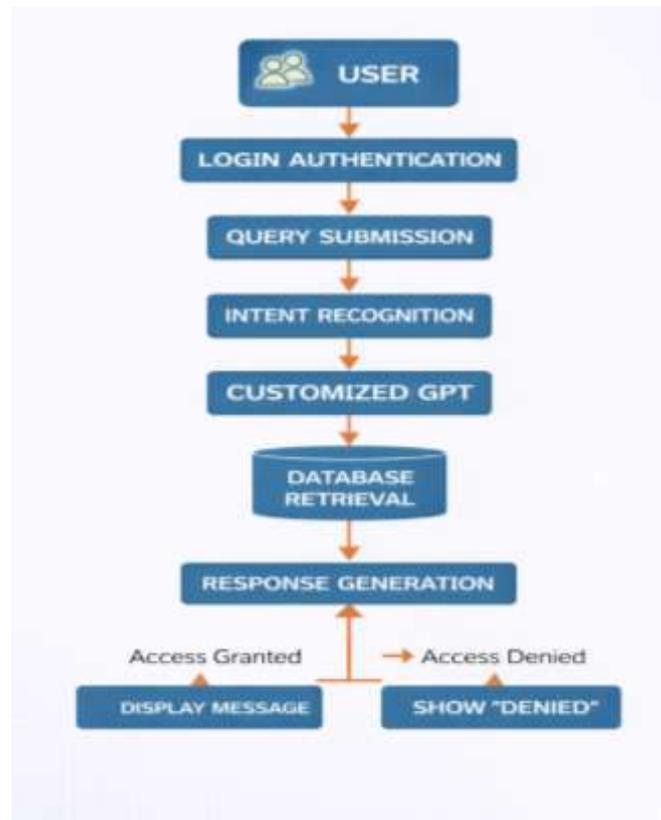• Generates context-aware responses

3.4 Database Layer

Stores:
• Academic information
• Placement records
• Institutional policies
• Student-related data

3.5 Architecture Flow Explanation

User → Login Authentication → Query Submission → Intent Recognition → Customized GPT → Database Retrieval → Response Generation → Output Display

IV. RESEARCH METHODOLOGY

The system was developed using a structured methodology involving data collection, model customization, system integration, and testing.

4.1 Data Collection and Preprocessing

Institutional documents, placement records, academic regulations, and FAQs were collected and preprocessed to remove inconsistencies and improve model understanding.

4.2 Model Customization

A pre-trained GPT model was fine-tuned using EGS-specific datasets to improve domain relevance and response accuracy.

4.3 Query Handling and Response Generation

User queries are processed using NLP techniques before being passed to the GPT model, ensuring meaningful and accurate responses.

4.4 System Integration and Testing

All modules were integrated and tested under different query scenarios to validate system reliability and performance.

V. RESULTS AND DISCUSSION

The customized GPT system was evaluated based on response accuracy, relevance, and user satisfaction.

5.1 Performance Evaluation

• Improved accuracy for EGS-related queries
• Reduced response time compared to manual systems
• High user satisfaction during testing

5.2 Comparative Analysis

Compared to rule-based chatbots, the proposed system demonstrates superior flexibility, scalability, and contextual understanding.

5.3 Observations

The system effectively handles complex natural language queries and provides consistent responses, making it suitable for institutional deployment.

VI. CONCLUSION

This paper presented the design and implementation of a customized college information chatbot using Natural Language Processing and Machine Learning techniques. The proposed system effectively automates the process of answering frequently asked queries related to admissions, courses, departments, examinations, and placement activities. By utilizing NLTK for text preprocessing and scikit-learn for intent classification, the chatbot is capable of understanding user queries and providing accurate responses in a conversational manner.

The system was developed as a lightweight and cost-effective solution that operates in a local environment without requiring continuous internet connectivity or high computational resources. The use of TF-IDF vectorization and supervised machine learning models enabled efficient intent detection and improved response accuracy. The modular architecture of the chatbot ensures ease of maintenance, scalability, and adaptability to different institutional requirements.

Overall, the proposed chatbot reduces the workload of administrative staff and enhances information accessibility for students and visitors. The results demonstrate that traditional NLP and machine learning techniques can be effectively applied to develop practical and reliable conversational systems for educational institutions. The system can be further enhanced in the future by integrating advanced deep learning models, web-based interfaces, and multilingual support to improve its intelligence and usability.

REFERENCES

[1]   A. Khanna and S. Kaur, "Building Intelligent Chatbots Using Natural Language Processing," *International Journal of Computer Applications*, vol. 178, no. 7, pp. 22–27, 2019.

[2]   M. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology," *Artificial Intelligence Applications and Innovations*, Springer, pp. 373–383, 2020.

[3]   T. K. Das and P. R. Sahoo, "Design and Implementation of an Educational Chatbot Using NLP," *International Journal of Engineering Research and Technology*, vol. 9, no. 6, pp. 112–116, 2020.

[4]   F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.