

Vivekanand Education Society's Institute of Technology



Department of Computer Engineering

Group No.: 06

Date :- 11/08/2024

Project Synopsis Template (2024-25) - Sem VII

SmartGuide: Chatbot for Efficient Employee Support and Document Analysis

Dr. Gresha Bhatia

Deputy HOD , Department of Computer Engineering

Shreya Hadkar

VESIT

2021.shreya.hadkar@ves.ac.in

Johan John

VESIT

2021.johan.john@ves.ac.in

Manali Patil

VESIT

2021.manali.patil@ves.ac.in

Vedang Rath

VESIT

2021.vedang.rathi@ves.ac.in

Abstract

SmartGuide is an advanced chatbot solution designed to enhance employee support and streamline organizational processes in large public sector organizations. Leveraging state-of-the-art deep learning and natural language processing techniques, SmartGuide efficiently addresses a wide range of queries related to HR policies, IT support, company events, and more. The chatbot integrates robust document processing capabilities, allowing it to analyze, summarize, and extract relevant information from user-uploaded documents.

Scalable to handle at least five simultaneous users, SmartGuide ensures prompt responses with a maximum latency of five seconds under normal conditions. Additionally, it incorporates two-factor authentication (2FA) for enhanced security and features a system-maintained dictionary for filtering inappropriate language. This comprehensive solution aims to improve operational efficiency and provide reliable, secure assistance to employees.

Introduction

In large public sector organizations, employees often face challenges in accessing accurate and timely information related to HR policies, IT support, company events, and other organizational matters. Traditional methods of obtaining this information can be time-consuming and inefficient, leading to delays and potential misunderstandings. Additionally, the need for secure handling of sensitive information and the ability to process and analyze documents further complicates the situation.

This project aims to develop an intelligent chatbot that leverages deep learning and natural language processing (NLP) techniques to address these challenges. The chatbot will serve as a centralized, easily accessible platform where employees can quickly and accurately get answers to their queries. By integrating document processing capabilities, the chatbot will also allow users to upload and analyze documents, extract relevant information and summarize content to meet organizational needs.

Furthermore, the chatbot is designed to be scalable, ensuring it can handle multiple users simultaneously without compromising response time. Enhanced security features, including two-factor authentication (2FA), will safeguard sensitive information, while a built-in bad language filter will maintain a professional environment. Ultimately, this project will streamline information access, improve efficiency, and enhance the overall employee experience within the organization.

Problem Statement

To develop a chatbot using deep learning and natural language processing (NLP) techniques to accurately understand and respond to queries from employees of a large public sector organization.

Features are as follows:

- Handling diverse questions related to HR policies, IT support, company events, and other organizational matters.
- Develop document processing capabilities for the chatbot to analyze and extract information from documents uploaded by employees. This includes summarizing a document or extracting text from documents relevant to organizational needs.
- Ensure the chatbot architecture is scalable to handle minimum 5 users parallelly. This includes optimizing response time (Response Time should not exceed 5 seconds for any query unless there is a technical issue like connectivity, etc.)
- Enable 2FA (2 Factor Authentication) in the chatbot for enhancing the security level of the chatbot.
- Chatbot should filter bad language as per system-maintained dictionary.

Proposed Solution

The chatbot will provide a centralized platform where employees can obtain accurate and timely information on various topics such as HR policies, IT support, and company events. The solution involves the following key components:

1. Query Handling

- **Natural Language Understanding:** The chatbot utilizes advanced NLP models, such as BERT or GPT, along with Dialogflow for robust intent recognition and entity extraction. Dialogflow processes user queries to identify intents and extract key information, which is then enhanced by the NLP models to provide accurate, context-aware responses to a wide range of employee inquiries related to HR policies, IT support, and company events.

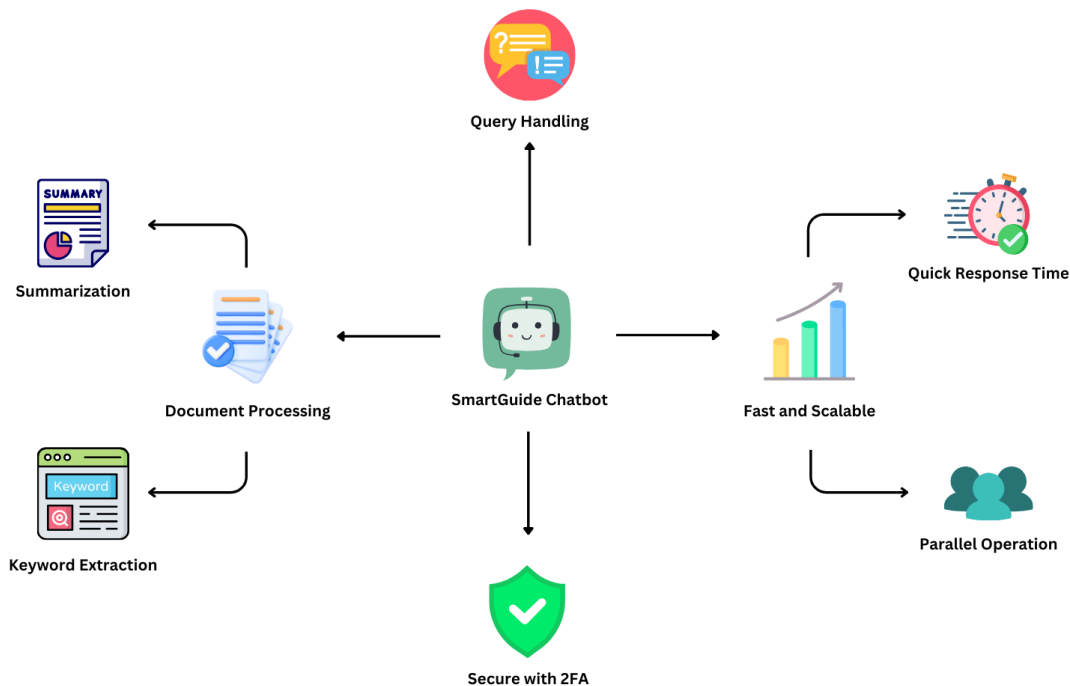
2. Document Processing

- **Text Extraction and Analysis:** The chatbot integrates document processing capabilities, including OCR for text extraction and NLP techniques for

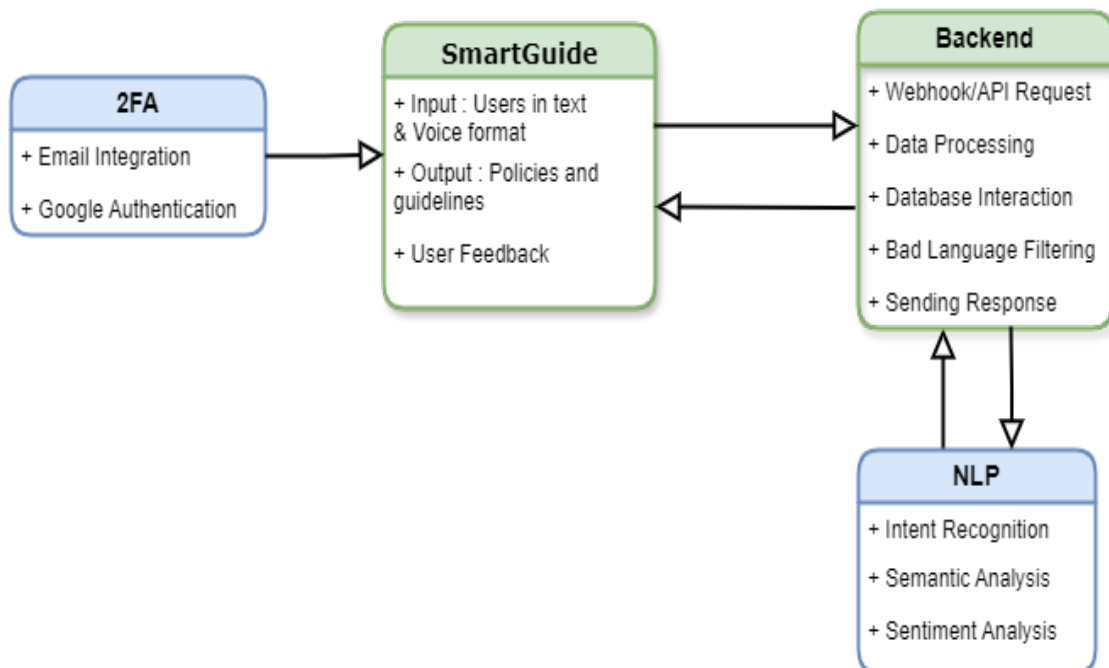
summarization and keyword extraction. This allows employees to upload documents and receive concise, relevant information or summaries based on the content.

3. Security and Filtering

- **Two-Factor Authentication (2FA):** To enhance security, the chatbot includes 2FA for verifying user identity, safeguarding sensitive information.
- **Bad Language Filtering:** A system-maintained dictionary and NLP-based models are employed to filter out inappropriate language, ensuring a professional and respectful interaction environment.



Block Diagram :



Methodology

1. Requirements Gathering

The first step in developing the chatbot involves understanding the specific needs of the public sector organization. This includes identifying the types of queries the chatbot must handle, such as HR policies, IT support, and company events, as well as the need for document processing capabilities. The security requirements, including two-factor authentication (2FA) and bad language filtering, are also crucial. This phase involves meetings with stakeholders, reviewing existing public sector documentation, and analyzing performance requirements.

2. System Design

Once the requirements are clear, the next step is to design the system architecture. The chatbot's architecture is divided into several components: frontend, NLP engine (Dialogflow), backend, and database.

- **Frontend Design:** The frontend is the user interface where employees will interact with the chatbot. It is essential to ensure the interface is user-friendly and responsive, capable of handling multiple users simultaneously.
- **NLP Engine Setup:** Dialogflow is selected as the NLP engine for its ability to recognize intents and entities, which allows the chatbot to understand and process user queries effectively.
- **Backend Development:** The backend is responsible for processing data, interacting with the database, and handling webhook requests from Dialogflow. It also manages document processing tasks like OCR (Optical Character Recognition), text extraction, and summarization.
- **Database Design:** The database stores user data, HR policies, IT support information, and documents. It is structured to facilitate quick retrieval of information based on user queries.
- **Security:** Two-factor authentication (2FA) is integrated to enhance security. Additionally, a bad language filter is implemented to maintain professional communication.

3. Development

The development phase involves building each component:

- **Frontend Development:** The chatbot interface is developed using modern JavaScript frameworks like React.js or Angular, ensuring it is intuitive and responsive.
- **Dialogflow Integration:** Intents and entities are configured in Dialogflow, allowing the chatbot to recognize and respond to different types of queries. Dialogflow is then integrated with the frontend to process user inputs.
- **Backend Development:** The backend is built using Python (Flask/Django) or Node.js. It handles Dialogflow webhook requests, processes documents using NLP models like BERT for summarization and keyword extraction, and implements API endpoints for 2FA and bad language filtering.
- **Database Setup:** A relational or NoSQL database (e.g., PostgreSQL, MongoDB) is set up to store and manage user data, documents, and organizational information.

4. Testing

Extensive testing is conducted to ensure the system operates as expected:

- Unit Testing: Individual components, such as the chatbot's response accuracy and document processing features, are tested.
- Integration Testing: The integration between frontend, backend, Dialogflow, and the database is tested to ensure seamless communication.
- Performance Testing: The system is tested to handle multiple users simultaneously, ensuring response times do not exceed 5 seconds.
- Security Testing: The 2FA mechanism and bad language filter are tested to verify their effectiveness.

5. Deployment

The chatbot is deployed to a cloud platform (e.g., AWS, GCP) or a dedicated server. Continuous Integration/Continuous Deployment (CI/CD) pipelines are set up to automate deployment and updates.

6. Maintenance and Iteration

Post-deployment, the system is continuously monitored, and user feedback is gathered. Performance metrics and feedback are used to make iterative improvements to the chatbot, ensuring it meets evolving organizational needs.

Hardware and Software Requirements

Hardware:

- Development Machine: Intel Core i5 with 8 GB RAM and 256 GB SSD
- Server (for Deployment): Multi-core CPU with 8 GB RAM

Software:

- Python: For backend development and integration.
- Dialogflow: For natural language understanding and chatbot framework.
- Flask or Django: For building the backend API and handling requests.
- TensorFlow or PyTorch: For custom deep learning models, if required.
- PDF Libraries: PyMuPDF or PyPDF2 for extracting text from PDF documents.
- Text Summarization: Libraries such as Hugging Face Transformers for summarization models.
- 2FA Integration: Use Google Authenticator for two-factor authentication.

Proposed Evaluation Measures

The assessment of the chatbot project will concentrate on multiple crucial areas in order to guarantee its efficiency and dependability. The measurement of accurate query responses will involve analyzing the percentage of correct answers in categories such as HR and IT, as well as using user satisfaction surveys to assess the relevance and clarity of the responses. The chatbot's performance will be evaluated by testing its capability to manage multiple users simultaneously within a response time less than 5 seconds. The efficiency of document processing will be assessed by comparing the accuracy of text extraction and summarization against manually verified data, as well as by monitoring the time required for processing these requests. Security measures, including two-factor authentication (2FA), will be thoroughly reviewed to identify and address any potential vulnerabilities. Finally, the effectiveness of the bad language filter will be assessed by evaluating its performance against a list of inappropriate words to ensure both accuracy and contextual relevance.

Conclusion

The chatbot is designed to address the intricate problems that workers in a sizable public sector organization face. Through the integration of comprehensive natural language processing and deep learning methodologies, along with a safe and expandable framework, the chatbot seeks to optimise information retrieval, augment efficiency, and elevate staff satisfaction in general. Its extensive features—which include real-time inquiry handling and document processing—are intended to speed up decision-making and communication, which will ultimately result in a more responsive and orderly workplace.

References

- [1] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018, October 11). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv.org. <https://arxiv.org/abs/1810.04805>
- [2] P. Kumar, M. Sharma, S. Rawat and T. Choudhury, "Designing and Developing a Chatbot Using Machine Learning," 2018 International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2018, pp. 87-91, doi: 10.1109/SYSMART.2018.8746972.
- [3] Lalwani, T., Bhalotia, S., Pal, A., Rathod, V., & Bisen, S. (2018). Implementation of a Chatbot System using AI and NLP. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3531782>
- [4] IRO Journals. (n.d.). Building Customized Chatbots for Document Summarization and Question Answering using Large Language Models using a Framework with OpenAI, Lang chain, and Streamlit - Research at York St John. <https://ray.yorks.ac.uk/id/eprint/9863/>
- [5] Prasad, Rajesh S., U. V. Kulkarni, and Jayashree R. Prasad. "Machine learning in evolving connectionist text summarizer." In 2009 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication.
- [6] H. Liang, "Research on Pre-training Model of Natural Language Processing Based on Recurrent Neural Network," 2021 IEEE 4th International Conference on Information Systems and Computer Aided Education (ICISCAE), Dalian, China, 2021.
- [7] S. Meshram, N. Naik, M. VR, T. More and S. Kharche, "Conversational AI: Chatbots," 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 2021.
- [8] Kooli, C. (2023). Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions. *Sustainability*, 15(7), 5614. <https://doi.org/10.3390/su15075614>
- [9] Khurana, Diksha & Koli, Aditya & Khatter, Kiran & Singh, Sukhdev. (2022). Natural Language Processing: State of The Art, Current Trends and Challenges. *Multimedia Tools and Applications*. 82. 10.1007/s11042-022-13428-4.