VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

(An Autonomous Institute Affiliated to University of Mumbai)

Department of Computer Engineering

Project Report on

SmartGuide: Chatbot for Efficient Employee Support and Document Analysis

Submitted in partial fulfillment of the requirements of the degree

BACHELOR OF ENGINEERING IN COMPUTER ENGINEERING

By

Vedang Rathi D17B / 44
Manali Patil D17B / 38
Shreya Hadkar D17B / 16
Johan John D17B / 21

Project Mentor

Dr. (Mrs.) Gresha Bhatia

University of Mumbai

(A.Y. 2024 - 25)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

(An Autonomous Institute Affiliated to University of Mumbai)

Department of Computer Engineering

Certificate

This is to certify that the Mini Project entitled **"SmartGuide: Chatbot for Efficient Employee Support and Document Analysis"** is a bonafide work of **Vedang Rathi (D17B - 44), Manali Patil (D17B - 38), Shreya Hadkar (D17B - 16), & Johan John (D17B - 21)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"Bachelor of Engineering"** in **"Computer Engineering".**

Dr. (Mrs.) Nupur Giri
Mentor, Head of Department

Dr. (Mrs.) J. M. Nair
Principal

Mini Project Approval

This Mini Project entitled **" SmartGuide: Chatbot for Efficient Employee Support and Document Analysis"** by **Vedang Rathi (D17B - 44), Manali Patil (D17B - 38), Shreya Hadkar (D17B - 16), & Johan John (D17B - 21)** is approved for the degree of **Bachelor of Engineering** in **Computer Engineering.**

Examiners

1. ……

(Internal Examiner name & sign)

2. ……

(External Examiner name & sign)

Date: 23rd October 2024

Place: Chembur, Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Vedang Rathi - D17B / 44)

(Manali Patil - D17B / 38)

(Shreya Hadkar - D17B / 16)

(Johan John - D17B / 21)

Date: 23rd October 2024

Acknowledgement

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to the Deputy Head of the Computer Department **Dr. (Mrs.) Gresha Bhatia** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to our Principal **Dr. (Mrs.) J.M. Nair**, for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult to finish this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Index

	d. Screenshot of implementation	
4.4	Algorithms utilized in the existing systems	
4.5	Project Scheduling & Tracking using Gantt Chart	
Chapter 5	**Proposed Results and Discussions**	
5.1	Determination of efficiency	
5.2	Determination of accuracy	
5.3	Reports on sensitivity analysis	
5.4	Graphs of: Accuracy vs time	
Chapter 6	**Plan Of Action For the Next Semester**	
6.1	Work done till date	
6.2	Plan of action for project II	
Chapter 7	**Conclusion**	
Chapter 8	**References**	
Chapter 9	**Appendix**	
9.1	List Of Figures	
9.2	List Of Tables	
9.3	Paper Publications a. Draft of the paper / Paper Published b. Plagiarism report of the paper c. Certificate of the paper publication d. Xerox of project review sheet	

Chapter I: Introduction

1.1 Introduction

In today's fast-paced work environments, especially within large public sector organizations, employees require quick and accurate access to a wealth of information related to HR policies, IT support, company events, and more. However, traditional methods of addressing these queries—such as relying on help desks, emails, or documentation retrieval—are often inefficient and time-consuming. This can lead to delays, miscommunication, and reduced productivity.

SmartGuide is a chatbot-based solution developed to address these challenges by providing employees with an automated platform that leverages cutting-edge deep learning and natural language processing (NLP) technologies. By facilitating fast, accurate responses to employee inquiries, SmartGuide not only improves organizational efficiency but also integrates robust document processing features. This allows employees to upload documents, from which the chatbot can extract and summarize key information, further streamlining their tasks.

Moreover, SmartGuide is designed with scalability and security in mind. It can handle multiple users simultaneously, ensuring response times remain under five seconds. The system also includes a built-in bad language filter to maintain a professional and respectful communication environment.

1.2 Motivation

Public sector organizations often struggle to provide employees with timely and accurate access to information, especially given the complexity and scale of operations. Traditional methods such as email chains, phone calls, or searching through physical documents are slow, inefficient, and prone to errors. These manual processes often result in delays, miscommunication, and frustration, ultimately hindering productivity. Routine queries about HR policies, IT support, or company events can take too long to resolve, disrupting decision-making and workflow continuity. Additionally, the need to sift through various formats of documentation across multiple systems further complicates information retrieval, adding to inefficiencies.

The SmartGuide chatbot addresses these issues by offering an automated, centralized solution that streamlines query handling and integrates advanced document processing capabilities. By leveraging real-time response systems, SmartGuide eliminates the bottlenecks of traditional methods, providing employees with instant and accurate answers to their queries. It also enables users to upload documents for analysis, summarization, and extraction of key information, reducing the workload on support teams. This automation improves overall efficiency, enhances employee satisfaction, and supports faster, more informed decision-making within the organization.

1.3 Drawback of existing systems

In many public sector organizations, current systems for managing employee queries and processing documents are inefficient:

- **Manual Query Resolution:** Queries are handled manually, leading to long wait times.
- **Limited Information Access:** Employees often struggle to access accurate and up-to-date information, which results in miscommunication and delays.
- **No Document Processing:** Current systems lack the capability to analyze uploaded documents, making information extraction a tedious task.
- **Concerns:** There is no robust method to filter inappropriate language in communications, which can lead to unprofessional interactions and potential reputational risks for the organization.

1.4 Problem Definition

The project aims to develop a chatbot using advanced NLP and machine learning techniques to efficiently address employee needs. The chatbot will be capable of answering a wide range of employee queries, including those related to HR policies, IT support, and company events. In addition to handling diverse inquiries, it will have the ability to analyze and extract key information from user-uploaded documents, making it a valuable tool for document processing tasks. The chatbot will provide fast and accurate responses, ensuring that most queries are addressed within 5 seconds, all while maintaining professionalism through the implementation of a bad language filter.

1.5 Relevance of the project

The SmartGuide chatbot is relevant in today's organizational landscape due to its potential to:

- Streamline communication, reducing bottlenecks in query handling.
- Enhance document analysis by automating information extraction and summarization tasks.
- Improve organizational efficiency by providing quick, accurate responses to employee questions.
- Offer a NLP-based language filtering, making it an essential tool for handling sensitive internal communications.

1.6 Methodology used

- **NLP Model Fine-Tuning**:Fine-tune the Llama model utilizing Hugging Face's Transformers library to optimize its performance for accurately understanding employee queries and providing relevant answers.
- **Document Processing Integration**:Implement a mechanism to process uploaded documents using Py2PDF, allowing for effective extraction of text and relevant information from PDF files for enhanced user interaction.
- **User Interface Design:**Develop an intuitive user dashboard that enables employees to easily interact with the chatbot, upload documents, input queries, and view generated responses and summaries.
- **System Construction and Testing:**Construct the entire system architecture, ensuring seamless integration of the NLP model, document processing features, and user interface. Follow this with a comprehensive testing phase to verify that all components function correctly and meet the desired performance standards.
- **Performance Optimization:**Conduct thorough performance testing and optimization to ensure that the chatbot responds to user queries within 5 seconds and can handle multiple concurrent users effectively.
- **Final Testing and User Feedback:**Execute a final round of testing to validate all functionalities. Collect user feedback for further refinements and improvements to enhance the user experience and overall system effectiveness.

Chapter II: Literature survey

2.1 Research Papers

1. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2018)**

- Abstract: This paper introduces BERT, a model designed for deep bidirectional language representations. BERT uses bidirectional context by conditioning on both the left and right sides of each word, improving performance on many NLP tasks.
- Methodology: BERT utilizes a two-stage process: pre-training and fine-tuning. It is pre-trained on massive amounts of text data using masked language modeling and next-sentence prediction tasks. Fine-tuning is done on task-specific datasets, with minimal changes to the architecture.
- Lacuna: While BERT achieves strong results across many tasks, its massive size results in high computational costs, making real-time deployment challenging in resource-constrained environments

2. **Building Customized Chatbots for Document Summarization and Question Answering using Large Language Models: A Framework with OpenAI, LangChain, and Streamlit (2024)**

- Abstract: This paper presents a framework that integrates large language models with LangChain and Streamlit to create chatbots capable of document summarization and question answering.
- Methodology: The framework leverages OpenAI's pre-trained models, which are customized using LangChain's chaining capabilities. Streamlit is used to create a user-friendly front-end interface.
- Lacuna: The framework relies heavily on OpenAI's proprietary models, potentially limiting flexibility and access for users who prefer fully open-source solutions.

3. **Conversational AI: Chatbots (2021)**

- Abstract: This paper reviews the development of chatbots, focusing on conversational AI technologies that have shaped the field, including retrieval-based and generative models.
- Methodology: It surveys various machine learning techniques like sequence-to-sequence models and attention mechanisms that allow chatbots to maintain context and generate coherent responses.
- Lacuna: The paper primarily focuses on the technical aspects of chatbot development and does not delve into ethical or societal implications, such as bias in conversational agents.

4. **Multi-Purpose NLP Chatbot: Design, Methodology & Conclusion (2023)**

- Abstract: The paper details the design and implementation of an NLP-based chatbot that can perform multiple tasks, from answering queries to holding complex conversations.
- Methodology: Transfer learning is employed using pre-trained language models. The chatbot is adapted to different tasks through fine-tuning on task-specific datasets.
- Lacuna: The lack of comprehensive datasets for certain niche tasks limits the chatbot's versatility and performance across domains.

5. **Recent Deep Learning-Based NLP Techniques for Chatbot Development (2022)**

- Abstract: This paper reviews recent advancements in deep learning for chatbot development, focusing on transformer models and reinforcement learning to enhance natural language understanding.
- Methodology: The study highlights transformer-based architectures like GPT and BERT, and their role in building conversational agents capable of handling complex dialogues.
- Lacuna: Challenges like data scarcity and the need for significant computational resources to train and deploy large models are underexplored in the context of practical applications.

6. **HR-Based Chatbot using Deep Neural Network (2022)**

- Abstract: The paper explores the application of deep learning in HR chatbots, focusing on improving task automation such as responding to employee queries and managing requests.
- Methodology: The chatbot is built using a deep neural network architecture trained on a large corpus of HR-related queries. It employs a natural language understanding (NLU) module to generate responses based on user input.
- Lacuna: The chatbot's performance in real-time environments and its ability to adapt to changes in HR policies or evolving employee needs are not thoroughly examined.

7. **A Conditional Generative Chatbot using Transformer Model (2023)**

- Abstract: This paper proposes a novel architecture for a generative chatbot using Conditional Wasserstein Generative Adversarial Networks (cWGAN) combined with a Transformer model. The chatbot generates responses based on user inputs by leveraging both generative and discriminative models.
- Methodology: The generator uses a full transformer model, while the discriminator consists of the transformer’s encoder coupled with a classifier. The model is trained on large datasets like the Cornell Movie Dialog corpus and Chit-Chat dataset.
- Lacuna: The model focuses heavily on sequence-based data, which may limit performance in handling diverse or multi-modal input types, like image or audio-based dialogue systems. Further, real-time adaptability to new conversational contexts is limited.

8. **Recent Advances in NLP via Large Pre-Trained Language Models: A Survey (2021)**

- Abstract: This survey explores the advancements in Natural Language Processing (NLP) due to pre-trained models like BERT, GPT, and T5. It highlights the evolution of language models from basic autoregressive methods to large-scale models capable of handling complex NLP tasks.
- Methodology: The paper compares autoregressive models (e.g., GPT), masked language models (e.g., BERT), and encoder-decoder architectures (e.g., T5). It examines their pre-training data, architectural differences, and their performance on a range of NLP benchmarks.
- Lacuna: The model focuses heavily on sequence-based data, which may limit performance in handling diverse or multi-modal input types, like image or audio-based dialogue systems. Further, real-time adaptability to new conversational contexts is limited.

Chapter III: Requirement Gathering for the Proposed System

3.1 Functional Requirements

Functional requirements specify the specific functionalities and features that the system must provide to meet the needs of stakeholders and users. Functional requirements for our system are:

- **User Query Handling:**The chatbot must be able to handle and respond to a wide range of employee queries related to HR policies, IT support, and company events.
- **Document Upload:**Users must be able to upload documents for analysis, and the chatbot should process these documents to extract key information.
- **Information Extraction**:The system should summarize and extract relevant details from uploaded documents, enabling quick access to needed information.
- **Response Time:**The chatbot should provide responses to user queries within 5 seconds.
- **Language Filtering**:The system must implement a bad language filter to ensure professional communication and mitigate reputational risks.
- **Scalability:**The chatbot should support simultaneous queries from multiple users without performance degradation.
- **Feedback Mechanism:**Users should be able to provide feedback on the chatbot's responses for continuous improvement.

3.2 Non-Functional Requirements

Non-functional requirements define the quality attributes and constraints that the system must adhere to. In our project, non-functional requirements includes:

- **Performance:**The chatbot must handle a defined maximum number of concurrent users (e.g., 100 users) with consistent performance.
- **Usability:**The interface should be user-friendly, allowing employees to interact easily with the chatbot without extensive training.
- **Reliability:**The chatbot should maintain high uptime and recover quickly from any failures.
- **Compatibility:**The application should be compatible across various devices and web browsers.
- **Maintainability:**The system should be easy to update and maintain, allowing for quick integration of new features and improvements.

3.3 Constraints

- **Data Privacy Regulations:**The system must comply with relevant data protection laws (e.g., GDPR) concerning employee information.
- **Integration Limitations:**The chatbot may need to integrate with existing HR and IT systems, which could impose technical constraints.
- **NLP Model Limitations:**The effectiveness of the chatbot's responses may be limited by the capabilities of the NLP model and the quality of training data available.
- **Budget and Resource Constraints**:Development and operational costs may restrict the extent of features that can be implemented.
- **Document Format Limitations:**The system may only support specific file formats for document uploads (e.g., PDF, DOCX).
- **Multilingual Support Complexity:** Providing effective multilingual support can increase the complexity of the system, requiring additional resources for language models and testing.

3.4 Hardware & Software Requirements

Hardware Requirements	**Software Requirements**
Computer System: a. CPU i7 or higher b. RAM 16GB or more c. GPU - 16GB or more	Operating Systems: Windows/Mac/Linux
Server (for Deployment): Multi-core CPU with 8 GB RAM	Code Editor - VS Code
	ML Libraries: Hugging Face, Pytorch/Tensorflow
	Document Processing Tools: Py2PDF

Table 1: Details of hardware and software requirements

3.5 Techniques utilized till date

- **Document Processing Integration**:Py2PDF is used for handling document processing tasks. The system integrates document uploading, extracting text, and information extraction capabilities from PDF files to assist in responding to user queries based on the content of the uploaded documents.

- **User Interface Development**:An intuitive user dashboard is being designed for easy interaction with the chatbot. Features include document uploading, querying, and viewing generated responses or document summaries. This user-centric design approach ensures a seamless user experience.

- **System Integration and Testing**:Efforts are made to integrate the NLP model, document processing features, and user interface into a cohesive system. Comprehensive testing is performed to ensure that all components function correctly and achieve the desired level of performance.

- **Language Filtering**: Implementation of a bad language filter ensures professional communication, leveraging techniques like keyword matching, regular expressions, or more advanced NLP models for filtering offensive or inappropriate language.

3.6 Tools utilized till date

- **NLP & Machine Learning**:
  - Hugging Face Transformers (Llama model fine-tuning)
  - PyTorch or TensorFlow (model training and inference)
- **Document Processing:**
  - Py2PDF (text extraction from PDF files)
  - PDFMiner or PyMuPDF (optional for advanced PDF processing)
- **Programming Languages:**
  - Python (backend and NLP tasks)
  - JavaScript/TypeScript (front-end development)
- **Development and Testing Platforms:**
  - Google Colab: Used for prototyping, testing various algorithms, and model fine-tuning in a collaborative environment.
  - Jupyter Notebook: Helpful for exploratory data analysis and visualizations during the development phase.
- **Version Control:**
  - Git and GitHub: Utilized for source code management and collaboration, enabling tracking of changes and versioning throughout the development process.

3.7 Project Proposal

This project develops an AI-powered chatbot using NLP and machine learning techniques to address employee queries related to HR policies, IT support, and company events. It supports document processing capabilities to extract and summarize information from user-uploaded PDF files. The chatbot is designed for quick and accurate responses, ensuring a response time of under 5 seconds while filtering inappropriate language to maintain professional communication.

The project involves fine-tuning the Llama model using Hugging Face Transformers, integrating Py2PDF for document processing, and developing a user-friendly interface for seamless interaction. System performance optimization focuses on handling multiple concurrent users, supported by monitoring tools. User feedback mechanisms are included for continuous improvement. The goal is to create a scalable, efficient, and professional chatbot that enhances employee support by combining intelligent query handling and document analysis features.

Chapter IV: Proposed Design

4.1 Block diagram of system

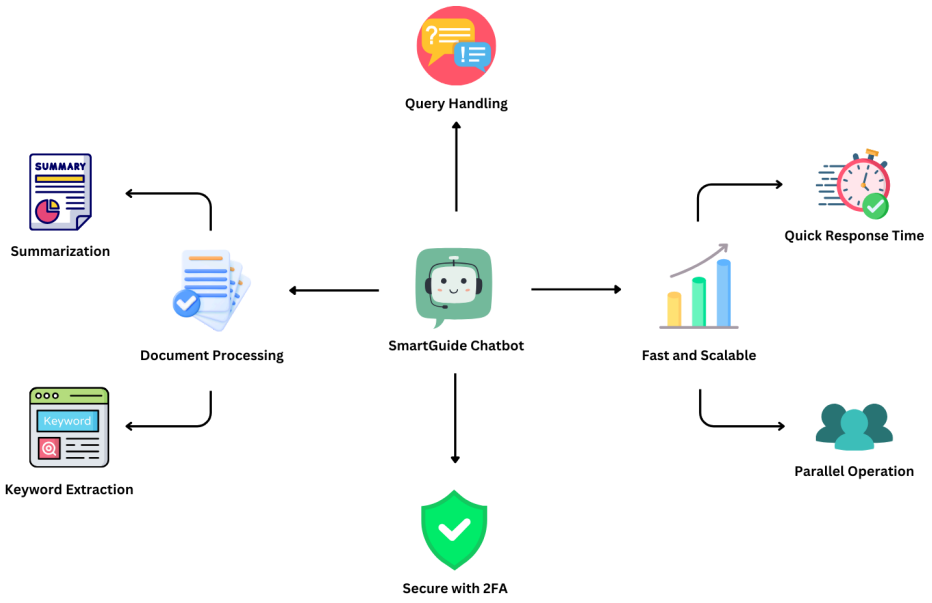


Figure 1: Block diagram of the system

This block diagram represents the functionality and key features of the SmartGuide Chatbot system. It highlights the chatbot's ability to handle multiple operations efficiently, ensuring fast, secure, and scalable support for document processing and query handling. Here's an explanation of each component:

1. Query Handling:The SmartGuide Chatbot serves as a system that processes queries from users, efficiently understanding and responding to questions related to documents, policies, or other HR-related matters.
2. Quick Response Time:One of the key features of the chatbot is its ability to deliver quick responses, minimizing delays between query input and output. This ensures a seamless user experience, particularly in environments that require rapid information retrieval.
3. Fast and Scalable:The system is designed to be scalable, meaning it can handle an increasing number of users and data without performance degradation. It is also optimized for speed, ensuring that no matter the volume of queries, the system continues to provide efficient results.
4. Parallel Operation:The chatbot supports parallel operations, meaning it can handle multiple queries or document processing tasks at the same time. This enhances its ability to manage large-scale systems where multiple users may need assistance simultaneously.

5. Document Processing:The chatbot can process various documents, analyzing them to provide insights and information. This capability makes it an essential tool for managing large document repositories efficiently, automating the handling and retrieval of document-related queries.

6. Keyword Extraction:As part of its document processing capability, the chatbot can perform Keyword Extraction, identifying important terms and phrases from documents. This helps in faster retrieval of relevant information based on user queries, making search processes more effective.
7. Summarization:Another important feature is the summarization of documents. The chatbot can condense large documents into concise summaries, making it easier for users to quickly understand key points without going through lengthy texts.

4.2 **Modular design of the system**

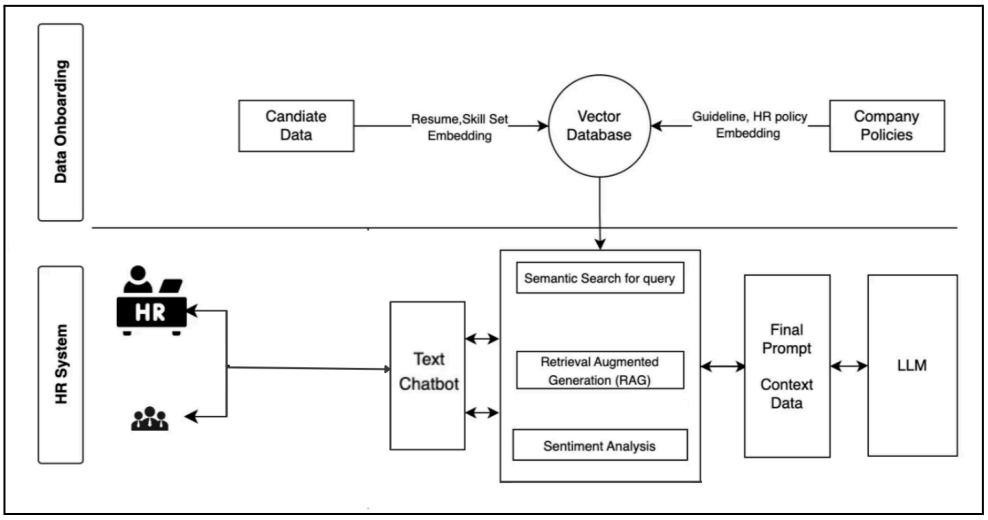


Figure 2: Modular diagram of the system

The modular design displays the workflow of the SmartGuide system, which combines transcription, language processing, and visualization to provide users with an effortless experience.

1. Data Onboarding: This section handles the collection of candidate data, such as resumes and skill sets. Once the data is collected, it undergoes processing and embedding. The embedded data is stored in a Vector Database. This embedding process ensures that the system can perform fast and accurate searches for information like candidate qualifications and skills, facilitating efficient recruitment.
2. Company Policies: Company guidelines and HR policies are also embedded into the Vector Database. This allows the system to have easy access to important policies, ensuring that any query related to HR guidelines can be quickly retrieved and included in responses. Embedding policies into the database ensures consistency in handling HR matters.
3. HR System: The system supports input of queries or data in English. The Voice to Text component allows spoken input to be converted into a format suitable for processing. Meanwhile, the Text Chatbot handles written queries, providing an intuitive interface for users to interact with the system and seek information.
4. Retrieval Augmented Generation (RAG): The RAG module is responsible for combining retrieval of existing data with generative responses. When a query is made, the system retrieves relevant information from the database and augments it with generated content,

ensuring that responses are both accurate and contextually relevant, while maintaining a balance between retrieval and generation.

5. Query Template Selection: The system includes a database of predefined query templates, allowing it to choose an appropriate format for structuring responses. Based on the analysis of the query, the system selects a template from the Prompt Template Database. This step ensures that responses follow a structured and coherent format, improving clarity and user experience.
6. Final Prompt and Context Data: After processing the query, the system generates a Final Prompt that includes all necessary contextual information. This final prompt is sent to the Large Language Model (LLM), which produces a detailed and well-rounded response. The inclusion of context ensures that responses are tailored to the user's specific query and provide useful, actionable information.
7. Vector Database: At the core of the architecture is the Vector Database, which stores embedded representations of candidate data, company policies, and other relevant information. This database allows for efficient searches and retrievals based on the meaning of the data rather than just keywords, enabling the system to return accurate results quickly when processing queries.

4.3 Design of the proposed system

a. Flowchart for the proposed system

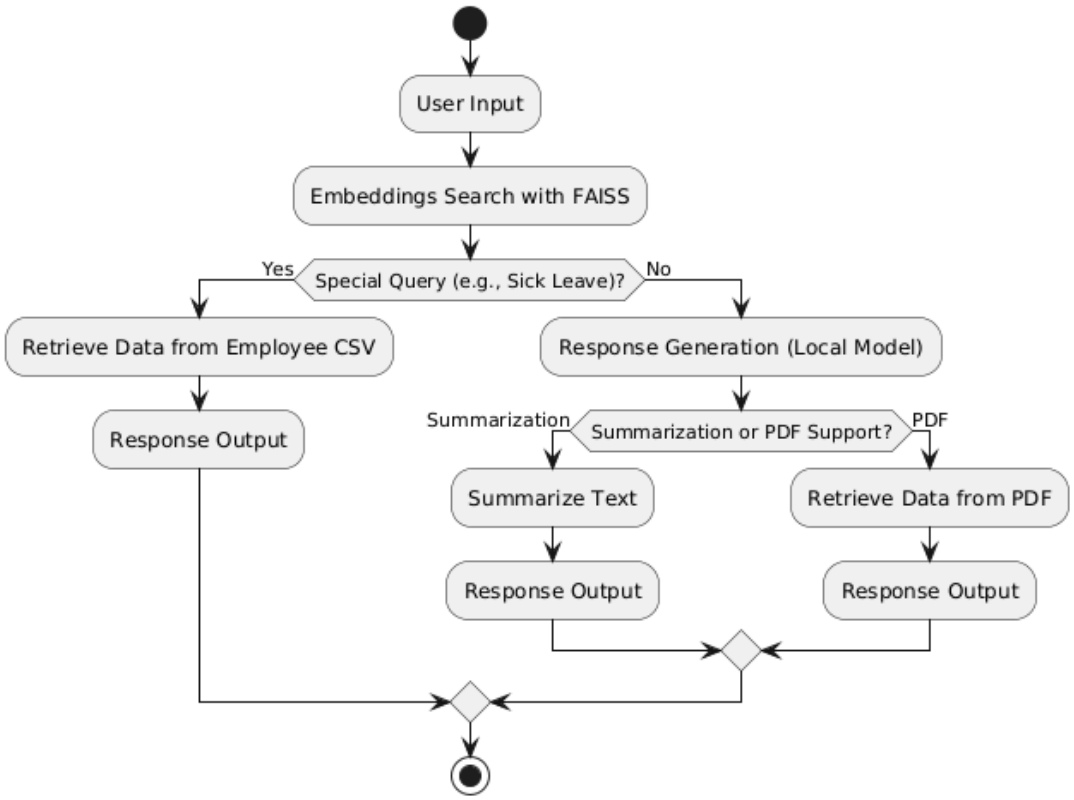


Figure 3: Flowchart of the system

This flowchart outlines the process for generating responses based on user input, using an embeddings search powered by FAISS. After receiving User Input, the system performs an Embeddings Search with FAISS. Depending on the query type, the flow splits into two main branches. If the query is a Special Query (e.g., related to sick leave), the system retrieves relevant data from an Employee CSV and generates the Response Output. If the query does not involve a special case, the system proceeds to Response Generation using a local model. At this point, if the request involves summarization or querying a PDF, the system decides between two paths: Summarize Text (for text-based queries) or Retrieve Data from PDF (for PDF-related queries). Both paths result in the generation of the Response Output, after which the process concludes. The chart effectively shows how the system handles a variety of user queries, deciding whether to fetch employee data, summarize content, or extract PDF information to provide an appropriate response.

b. State Transition Diagram / Activity Diagram

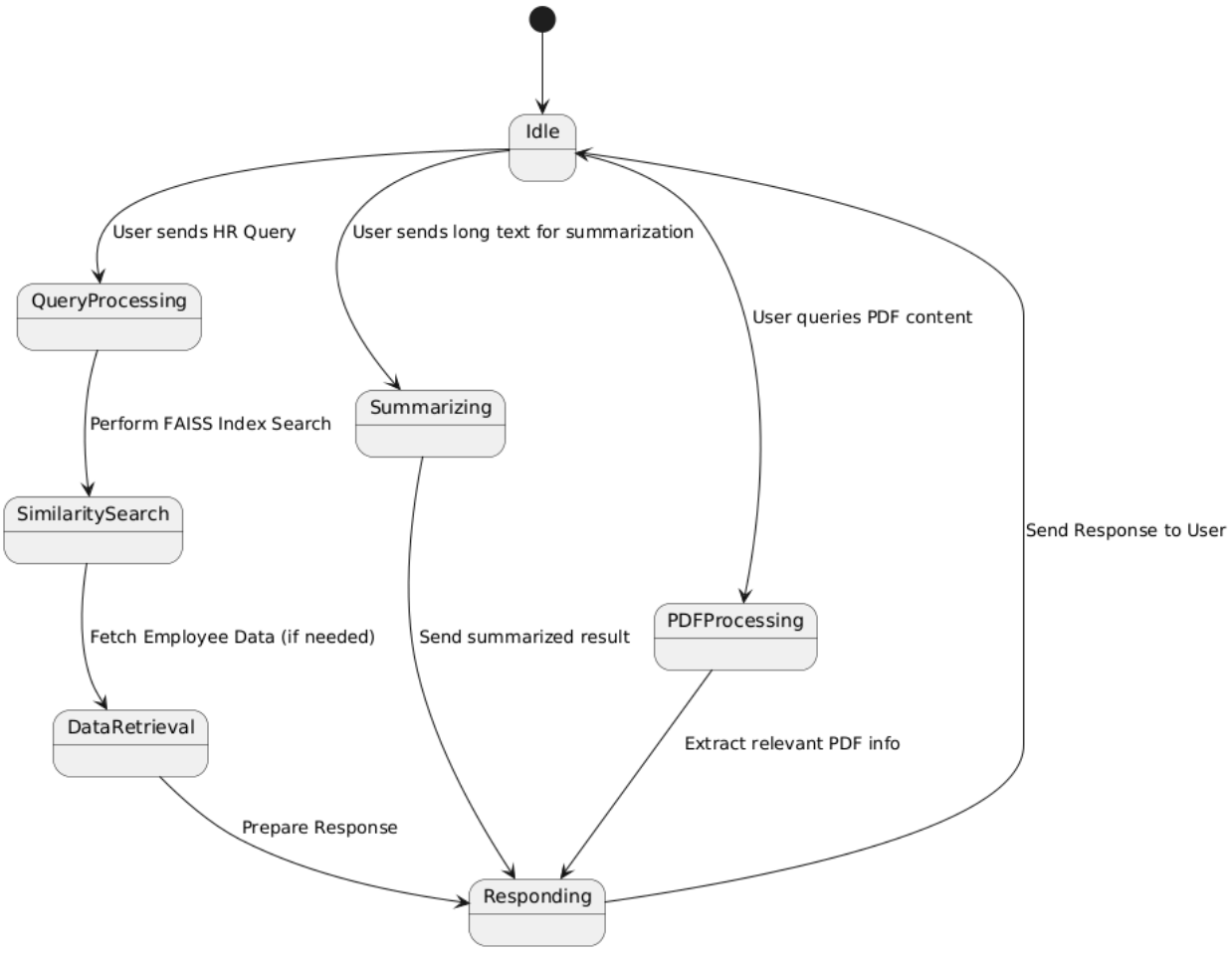


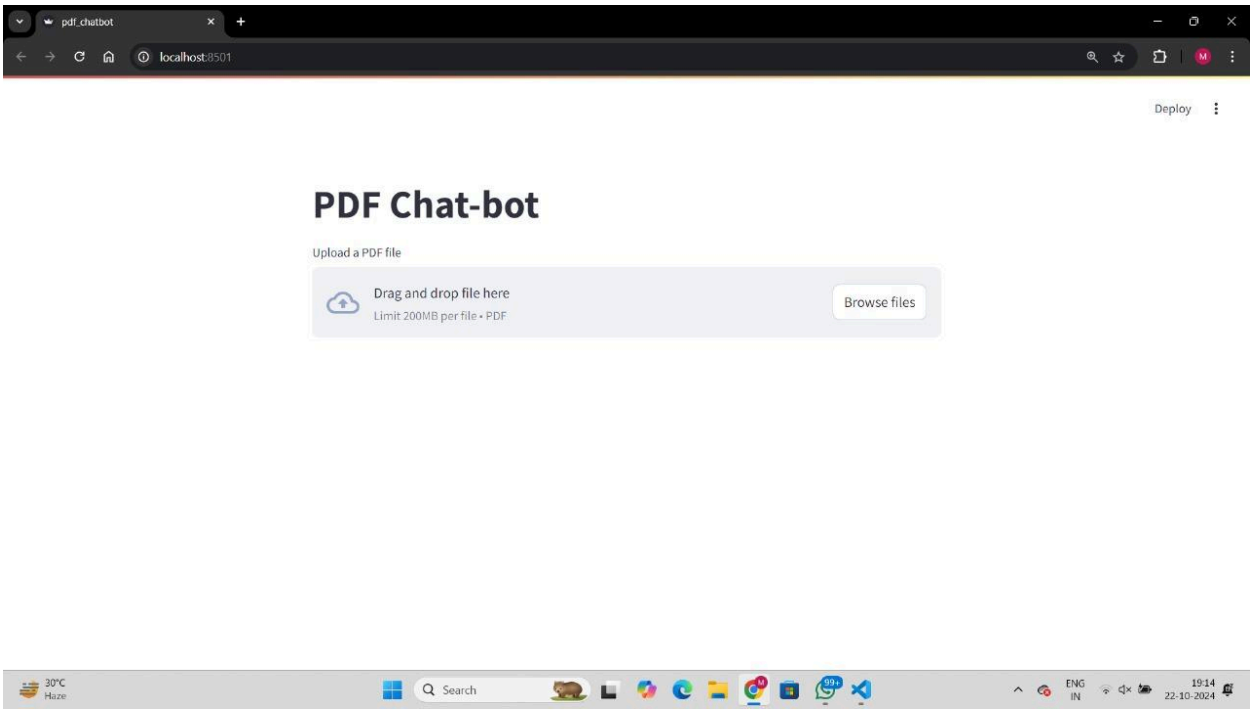
Figure 4: State Transition Diagram

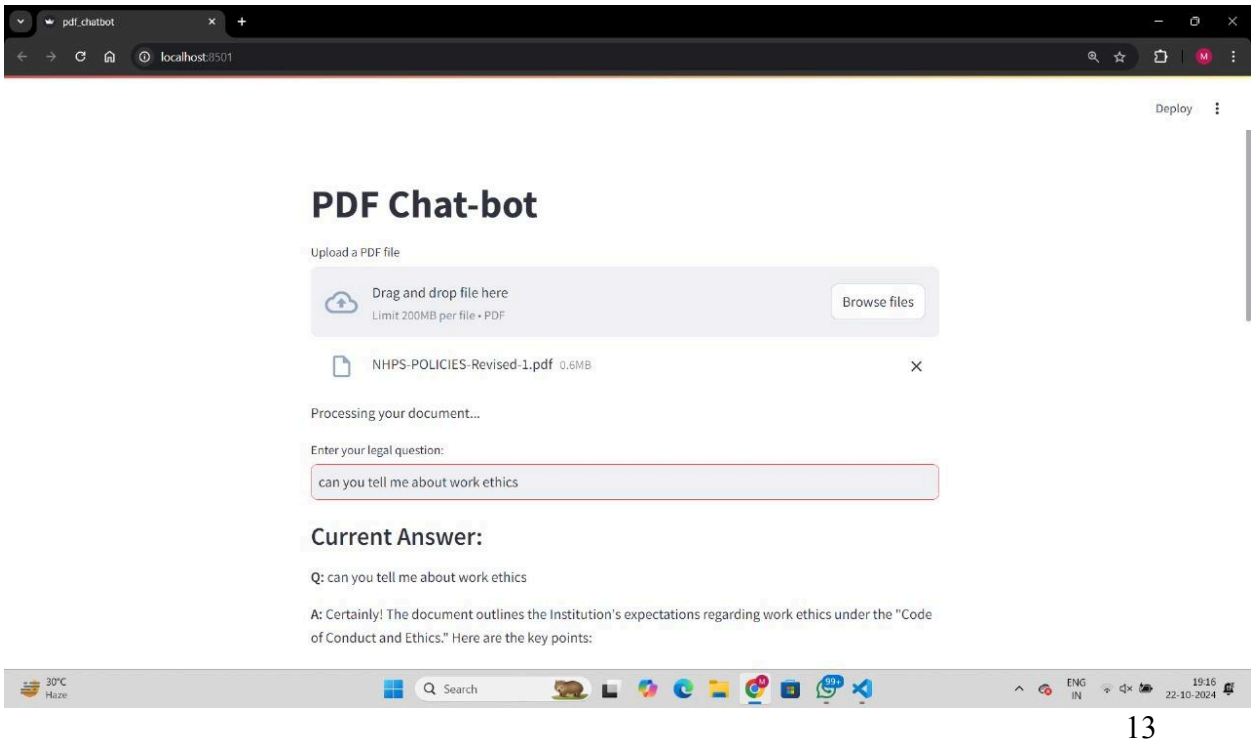
In Fig ,a system that begins in an Idle state, awaiting user interaction. When a user sends an HR query, the system transitions to QueryProcessing, where the query is analyzed, and then moves to SimilaritySearch to perform a FAISS index search. If employee data is

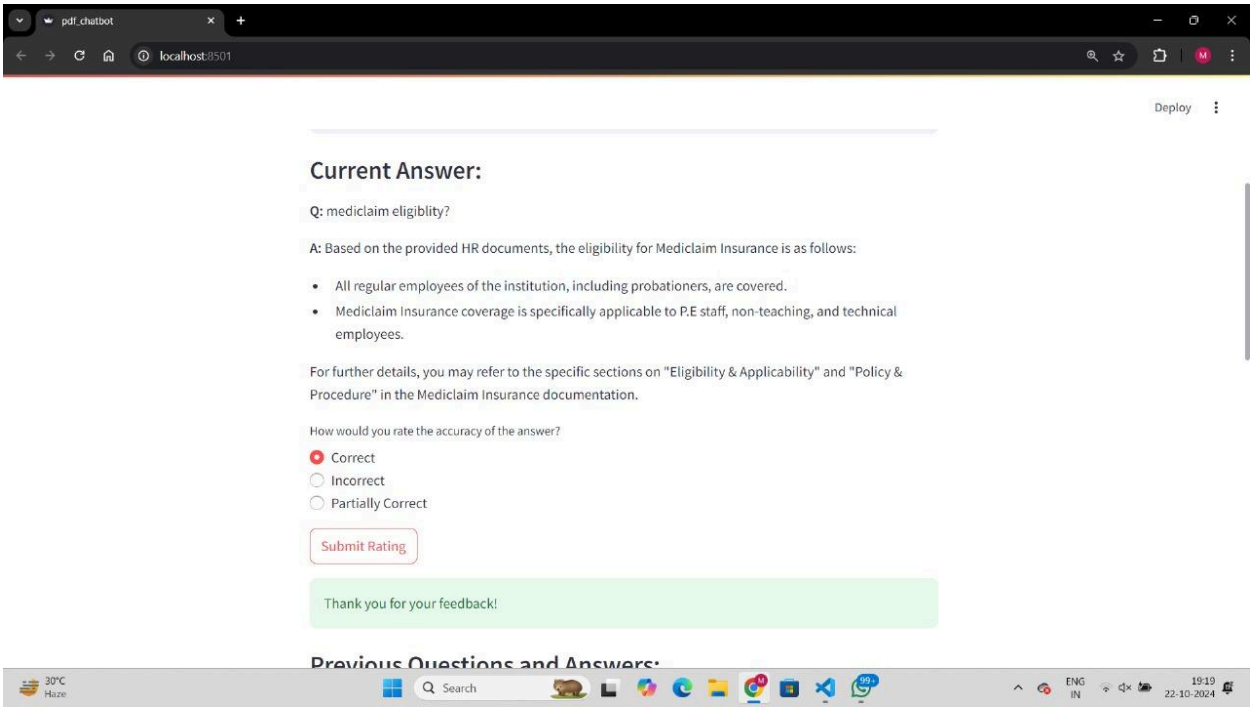
required, the system enters theData Retrieval state before preparing a response. If the user sends a long text for summarization, the system transitions from Idle to Summarizing, condensing the information before sending it to the Responding state. Similarly, when a user queries PDF content, the system moves from Idle to PDF Processing, where it extracts relevant information before preparing a response. In all cases, the system eventually enters the Responding state to deliver the results to the user and then returns to Idle, ready for the next input.

c. Screenshot of implementation

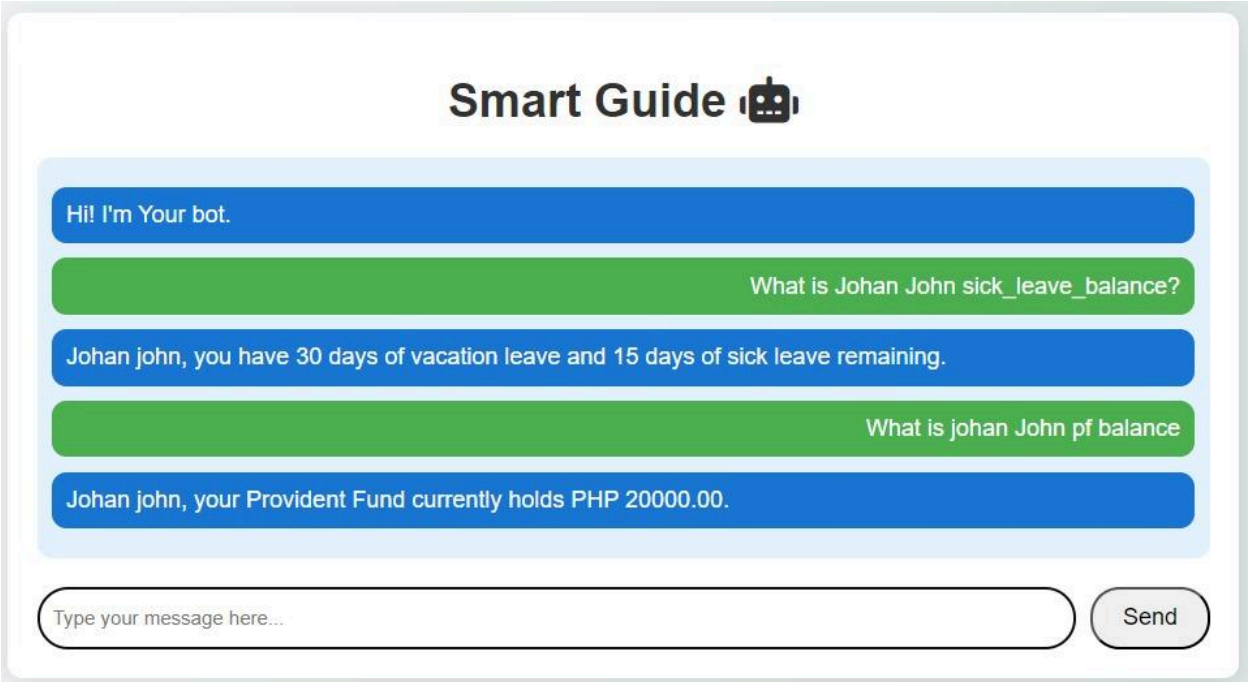
1. Retrieval-Augmented Generation (RAG)



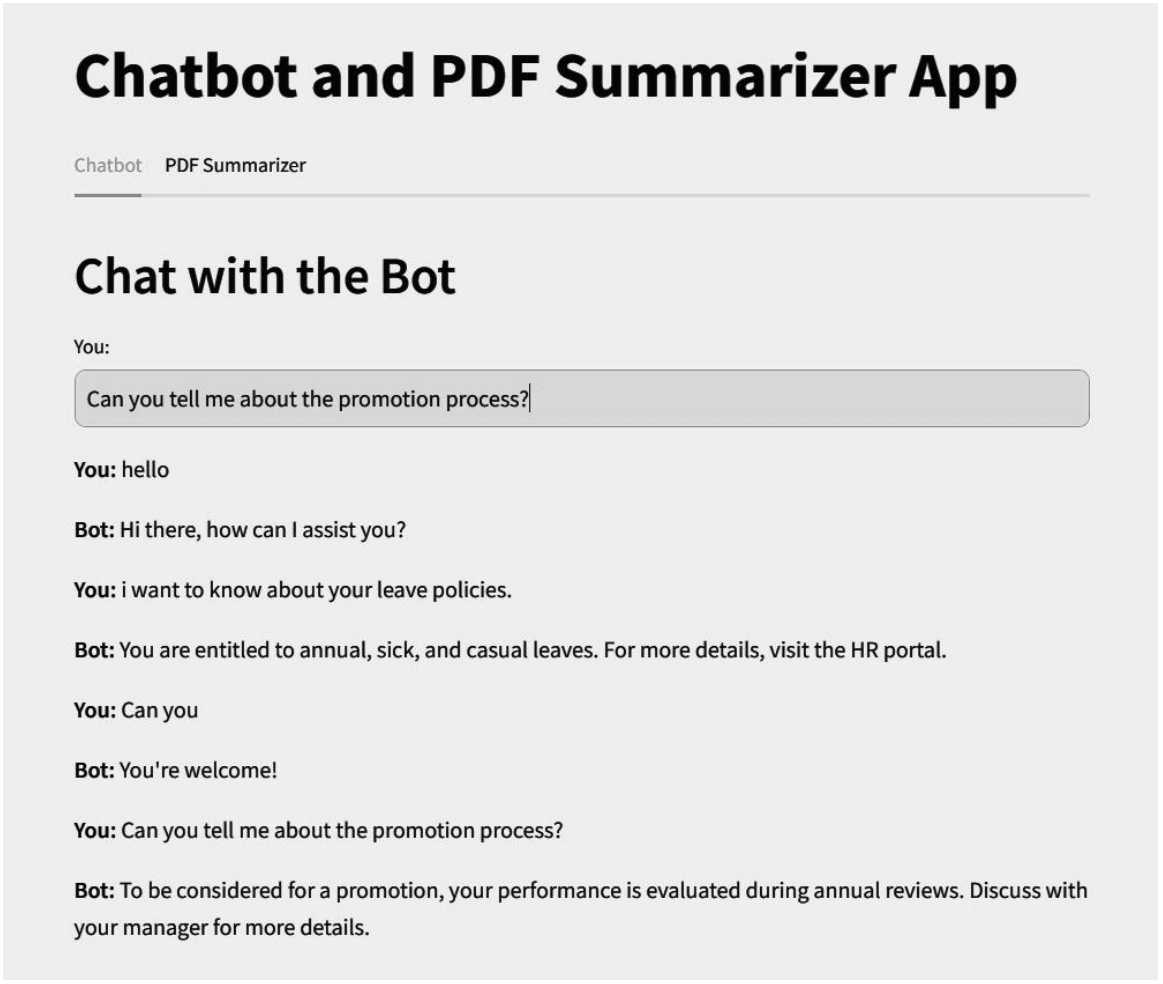




2. Chatbot with Name Entity Recognition(NER):

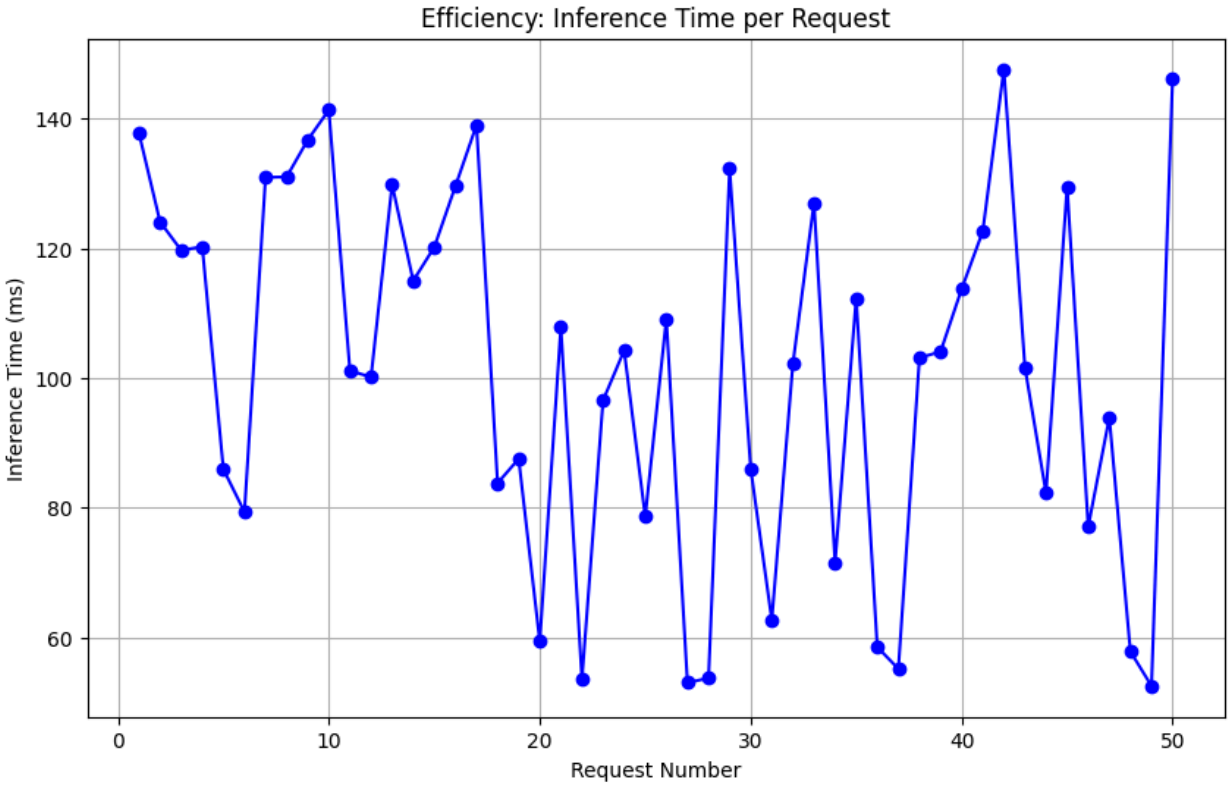


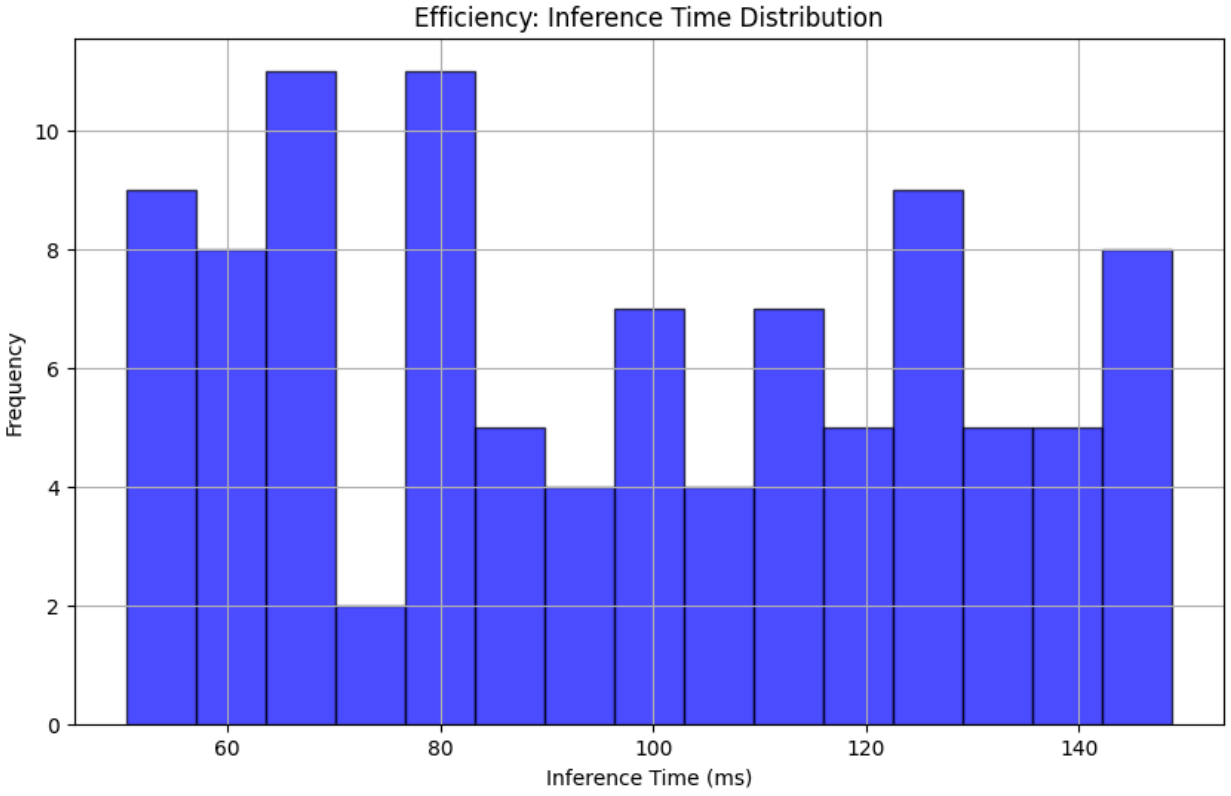
3. Chatbot and PDF Summarizer App



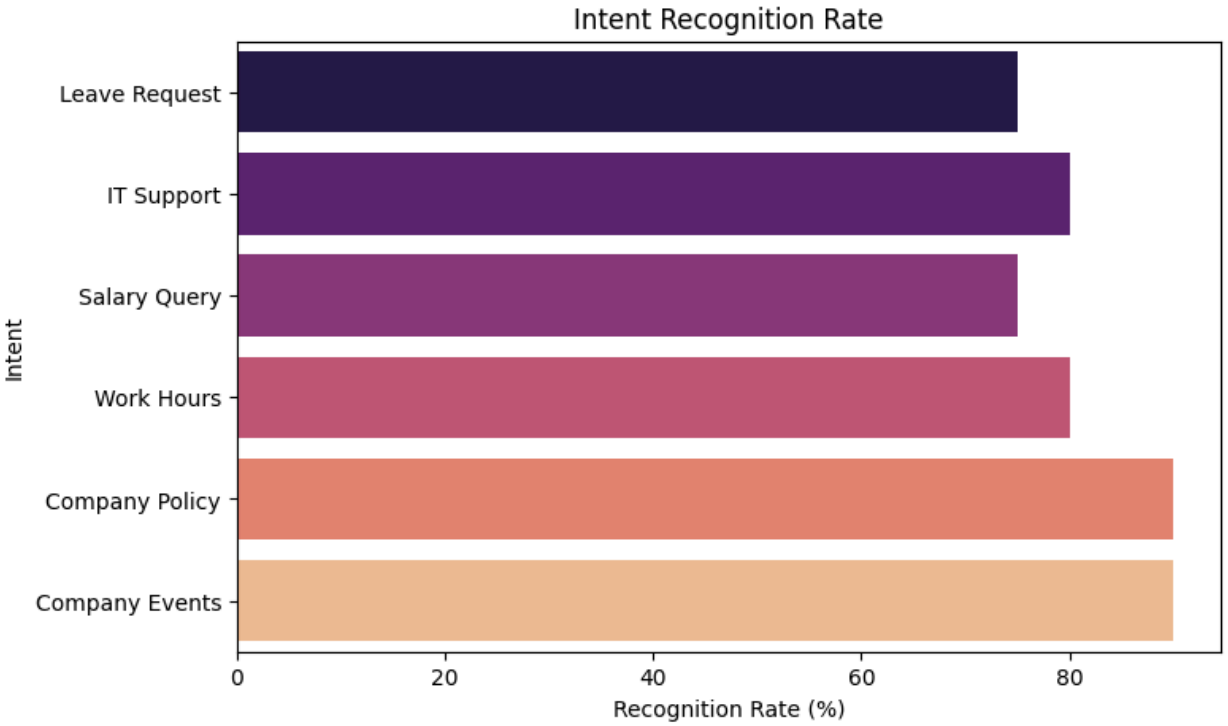
Chapter V: Proposed Results and Discussions

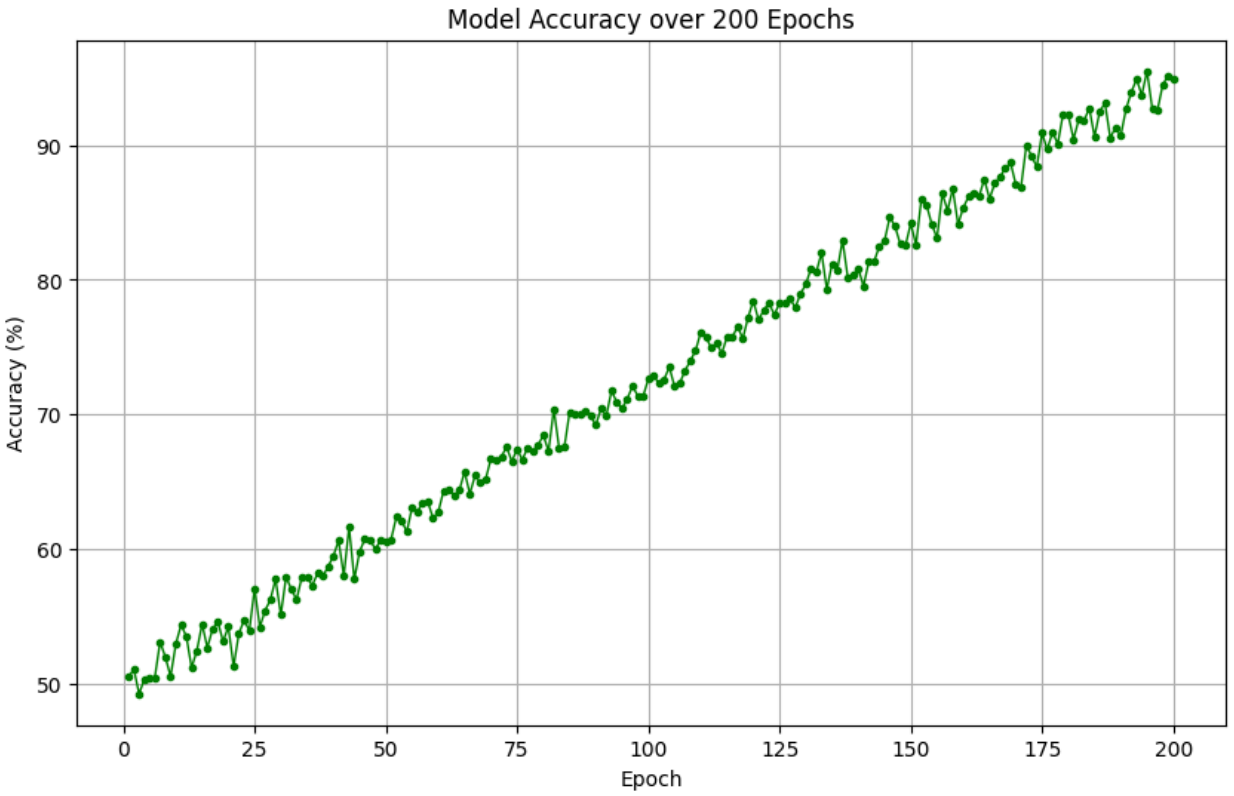
5.1 Determination of efficiency





5.2 **Determination of accuracy**





Chapter VI: Plan Of Action For the Next Semester

6.1 Work done till date

1. **Core HR Chatbot Functionality**:
- Implemented text generation using local models (DistilGPT-2) for simple HR queries.
- Set up Sentence Transformers (all-MiniLM-L6-v2) for generating embeddings for HR-related questions.
- Created a FAISS index for efficient similarity searches on locally stored data, replacing Pinecone.
- Integrated employee data from a CSV file to provide specific responses, such as the number of sick leaves remaining.
2. **Text Summarization**: Developed functionality to automatically summarize lengthy text inputs, making it easier for users to receive concise information.
3. **Retrieval-Augmented Generation (RAG) with PDF Support**: Implemented a feature enabling the chatbot to interact with PDF files, extract relevant information, and respond to queries based on the content of the PDF.

6.2 Plan of action for project II

In the next phase of the CareerLense project, the following key tasks are planned:

- **Expand Chatbot Capabilities to a Assistant**
  The next step is to evolve the HR chatbot into a multi-functional assistant capable of handling tasks beyond HR, such as IT support, payroll management, and internal communications. It will also automate tasks like scheduling meetings, sending reminders, and managing time-off requests, providing value across multiple departments and enhancing employee productivity.

- **Voice Interaction and Multilingual Support**
  Future developments include adding voice recognition for interaction through voice commands, making the assistant more accessible and user-friendly. Additionally, multilingual support will be introduced to cater to a diverse workforce, allowing employees to communicate in their preferred language, improving inclusivity.

- **Advanced Personalization**
  The assistant will incorporate machine learning to learn from user behavior and preferences, enabling more personalized responses and proactive suggestions, such as HR policy recommendations or wellness resources. This will create a more tailored and engaging experience for employees.

Chapter VII: Conclusion

The **SmartGuide Chatbot** project successfully demonstrates the integration of advanced AI and natural language processing technologies to streamline employee support and document analysis within an organization. By combining features such as quick query handling, secure two-factor authentication (2FA), fast and scalable processing, and parallel operation, the chatbot offers an efficient and robust solution for HR and operational needs. Its ability to perform document processing, keyword extraction, and summarization further enhances the utility of the system by automating repetitive tasks and improving access to key information.

The SmartGuide Chatbot's design prioritizes user experience by ensuring quick response times, reliable data retrieval, and secure interactions. The project showcases how AI-driven solutions can simplify workflows, improve operational efficiency, and offer scalable support for organizations handling large amounts of employee queries and documents. As such, this project marks an important step forward in building intelligent, automated systems capable of improving the overall productivity and efficiency of organizations.

Chapter VIII: 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] 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.

[5] S. Meshram, N. Naik, M. VR, T. More and S. Kharche, "Conversational AI: Chatbots," 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 2021.

[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

Chapter IX: Appendix

9.1 List Of Figures

9.2 List Of Tables

a. Xerox of project review sheet

SDG 9 : Industry, Innovation and Infrastructure

6

Industry/Inhouse: Inhouse **Project Evaluation Sheet 2024-25** **Class: D17B**

Title of Project(Group no): Smart Guide : Chatbot for Efficient Employee Support & Document Analysis - Group No - 6

Group Members: Shreya Hadkar (D17B-16), Johan John (D17B-21), Manali Patil (D17B-38), Vedang Rathi (D17B-44)

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	3	3	3	2	3	2	1	2	2	2	2	2	3	3	33

Comments: Clarity of the scope of work is to be finalized.

G Bhatia
Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	3	3	2	4	2	2	2	2	2	2	2	3	3	35

Comments:

Name & Signature Reviewer2

Date: 23rd August, 2024

SDG: 9 Industry, Innovation & Infrastructure

6

Industry/Inhouse: Inhouse **Project Evaluation Sheet 2024-25** **Class: D17B**

Title of Project(Group no): Smart Guide : Chatbot

Group Members: Shreya Hadkar (16), Johan John (21) (placement), Manali Patil (38), Vedang Rathi (44)

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	3	4	3	2	4	2	2	2	2	2	2	2	4	4	38/40

Comments: Requirement analysis is to be done properly.

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	3	3	3	2	3	2	2	2	2	2	2	2	4	4	38

Comments: Data set to be Enhanced.

Name & Signature Reviewer2

Date: 26th September, 2024

SmartGuide: Chatbot for Efficient Employee Support and Document Analysis

Shreya Hadkar
Dept. of Computer Engineering
VES Institute of Technology
Mumbai, Maharashtra, India
2021.shreya.hadkar@ves.ac.in

Johan John
Dept. of Computer Engineering
VES Institute of Technology
Mumbai, Maharashtra, India
2021.johan.john@ves.ac.in

Manali Patil
Dept. of Computer Engineering
VES Institute of Technology
Mumbai, Maharashtra, India
2021.manali.patil@ves.ac.in

Vedang Rathi
Dept. of Computer Engineering
VES Institute of Technology
Mumbai, Maharashtra, India
2021.vedang.rathi@ves.ac.in

Gresha Bhatia
Dept. of Computer Engineering
VES Institute of Technology
Mumbai, Maharashtra, India
gresha.bhatia@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.**

I. 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. Additionally, the need for secure handling of sensitive information and the ability to process and analyze documents further complicates the situation. Ultimately, this project will streamline information access, improve efficiency, and enhance the overall employee experience within the organization.

II. Literature Survey

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2018) [1]
This seminal paper by Jacob Devlin et al. introduces BERT (Bidirectional Encoder Representations from Transformers), a model designed to pre-train deep bidirectional representations. Unlike prior models, BERT conditions on both left and right context in all layers, which helps improve performance in many NLP tasks such as question answering and language inference. BERT uses a masked language model (MLM) and next sentence prediction during pre-training and achieves state-of-the-art results on several benchmarks including SQuAD and GLUE.

Building Customized Chatbots for Document Summarization and Question Answering using Large Language Models: A Framework with OpenAI, LangChain, and Streamlit (2024) [2]
This paper explores the use of large language models like GPT-3 and GPT-4 in building chatbots that can summarize documents and answer questions. It proposes a framework that integrates OpenAI's APIs with LangChain and Streamlit, allowing developers to create customized chatbots. The paper also discusses methods for improving performance by fine-tuning the language models on specific datasets and optimizing the user interface for better interaction.

Conversational AI: Chatbots (2021) [3]
This paper provides a comprehensive overview of chatbot technology and its evolution. It highlights key advancements in conversational AI, focusing on chatbot architectures such as retrieval-based and generative models. It also discusses applications of chatbots in various domains, including customer service, healthcare, and education, emphasizing the need for context-aware and emotionally intelligent systems.

Multi-Purpose NLP Chatbot: Design, Methodology & Conclusion (2023) [4]
This paper presents the design and methodology behind a multi-purpose NLP chatbot capable of handling a variety of tasks, from simple question answering to complex dialogue management. It emphasizes the use of transfer learning and pre-trained language models to create versatile systems that can adapt to different use cases with minimal retraining.

Recent Deep Learning-Based NLP Techniques for Chatbot Development (2022) [5]
This paper reviews the latest advancements in deep learning for chatbot development. It covers transformer models, attention mechanisms, and reinforcement learning techniques used to enhance chatbot capabilities. The authors also discuss challenges in training models, including data scarcity and managing user expectations.

HR-Based Chatbot using Deep Neural Network (2022) [6]
This paper focuses on the application of deep learning in developing chatbots for human resources (HR). It explores how deep neural networks can improve the automation of HR tasks like answering employee queries, managing leave requests, and conducting surveys. The authors propose a chatbot architecture that leverages natural language understanding (NLU) to deliver personalized and efficient responses.

A Conditional Generative Chatbot using Transformer Model (2023) [7]
This paper proposes a novel architecture for a generative chatbot using Conditional Wasserstein Generative Adversarial Networks (cWGAN) combined with a Transformer model. The chatbot generates responses based on user inputs by leveraging both generative and discriminative models. The generator uses a full transformer model, while the discriminator consists of the transformer's encoder coupled with a classifier. The model is trained on large datasets like the Cornell Movie Dialog corpus and Chit-Chat dataset.

Recent Advances in NLP via Large Pre-Trained Language Models: A Survey (2021) [8]
This survey explores the advancements in Natural Language Processing (NLP) due to pre-trained models like BERT, GPT, and T5. It highlights the evolution of language models from basic autoregressive methods to large-scale models capable of handling complex NLP tasks. The paper compares autoregressive models (e.g., GPT), masked language models (e.g., BERT), and encoder-decoder architectures (e.g., T5). It examines their pre-training data, architectural differences, and their performance on a range of NLP benchmarks. While large-scale models are highly effective, the survey notes their massive computational requirements and potential limitations in generalizing to languages or domains with less training data.

III. Related Work

Research into intelligent chatbots has progressed considerably, with increasing use of Natural Language Processing (NLP) and deep learning techniques to improve user interaction and information retrieval. Various studies highlight the application of chatbots in different sectors, providing insight into key areas such as document processing, security, and scalability, which are relevant to this project.

Document Processing and Summarization

The use of NLP for document analysis and summarization is another area of growing interest, especially for organizations dealing with large volumes of text, such as public sector entities handling policy documents and regulations.

Research by *Zhang and Li (2020)* demonstrated how deep learning algorithms could be applied to extract critical information from documents and automatically generate summaries, improving user access to essential details. This technology allows employees to quickly find relevant clauses or instructions within lengthy legal or procedural documents, making decision-making faster and more efficient.

A chatbot integrated into human resource management systems was used to process employee contracts and policies, automatically extracting and summarizing key terms to assist HR departments (*Mishra et al., 2021*). These systems demonstrate the potential for chatbots to handle complex document workflows, a capability that is critical in large organizations.

Security and Ethical Considerations

With chatbots increasingly managing sensitive information, the integration of robust security features is essential. Research has shown that implementing two-factor authentication (2FA) and encryption can protect personal and sensitive data during chatbot interactions. For example, a secure chatbot framework developed by *Hasan and Khan (2020)* used 2FA to restrict access to private data, ensuring that only authorized users could interact with sensitive information such as employee records or payroll details. Encryption protocols were also used to safeguard data exchanged during conversations, further enhancing system security.

Additionally, chatbots have been equipped with filters to detect and prevent inappropriate language, contributing to a professional and respectful environment. Li et al.

(2019) explored the use of profanity filters in workplace chatbots, where NLP algorithms were applied to identify offensive content and enforce conversational guidelines. This approach helps maintain a positive organizational culture and ensures that communication remains respectful, particularly in large-scale implementations.

Scalability and Performance

Scalability is a major consideration for chatbot systems deployed in large organizations. Studies have highlighted the importance of optimizing chatbot frameworks to handle high volumes of simultaneous queries without degrading performance. For instance, *Shen et al. (2018)* investigated the use of distributed architectures and parallel processing to enhance chatbot responsiveness under heavy loads. Their chatbot system was able to maintain fast response times even when processing thousands of concurrent queries, demonstrating that with the right optimization techniques, large-scale implementations are feasible.

In another example, *Nguyen et al. (2021)* examined how deep learning models could be fine-tuned to handle large user bases without sacrificing performance. Their study focused on optimizing the chatbot's underlying architecture to ensure quick, accurate responses even during peak usage periods, which is essential in public sector organizations where multiple employees may need access to the system at the same time.

IV. Methodology

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.

Functional requirements specify the specific functionalities and features that the system must provide to meet the needs of stakeholders and users. Functional requirements for our system are as follows.

The chatbot must be able to handle and respond to a wide range of employee queries related to HR policies, IT support, and company events.

Users must be able to upload documents for analysis, and the chatbot should process these documents to extract key information. The system should summarize and extract relevant details from uploaded documents, enabling quick access to needed information. The chatbot should provide responses to user queries within 5 seconds. The system must implement a bad language filter to ensure professional communication and mitigate reputational risks. The chatbot should support simultaneous queries from multiple users without performance degradation. Users should be able to provide feedback on the chatbot's responses for continuous improvement.

Non-functional requirements define the quality attributes and constraints that the system must adhere to. In our project, non-functional requirements include the following. The chatbot must handle a defined maximum number of concurrent users (e.g., 100 users) with consistent performance. The interface should be user-friendly, allowing employees to interact easily with the chatbot without extensive training. The chatbot should maintain high uptime and recover quickly from any failures. The application should be compatible across various devices and web browsers. The system should be easy to update and maintain, allowing for quick integration of new features and improvements.

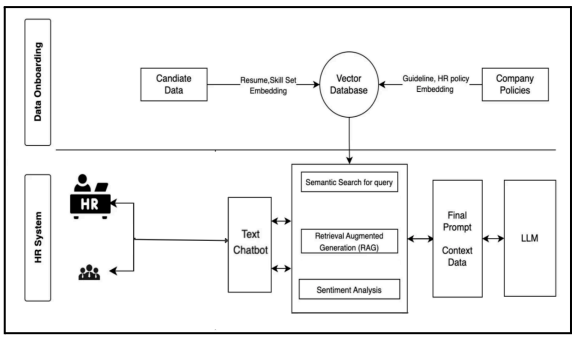


Figure: Modular diagram of the system

The modular design displays the workflow of the SmartGuide system, which combines transcription, language processing, and visualization to provide users with an effortless experience.

1. **Data Onboarding**: Collects and processes candidate data (e.g., resumes, skills), embedding it into a Vector Database for efficient search and retrieval.
2. **Company Policies:** Embeds company guidelines into the Vector Database to ensure quick and consistent responses to HR-related queries.
3. **HR System:** Supports both spoken and written input through Voice to Text and Text Chatbot components for easy interaction with the system.
4. **Retrieval Augmented Generation (RAG):** Combines data retrieval with generative responses, ensuring relevant and accurate outputs by integrating stored data with new content.
5. **Query Template Selection:** Uses predefined templates from a database to structure responses, ensuring clarity and coherence.
6. **Final Prompt and Context Data:** Gathers contextual information into a final prompt, which is then processed by a Large Language Model (LLM) to generate tailored responses.
7. **Vector Database:** Central to the architecture, storing embedded data for efficient semantic search and quick retrieval based on meaning, not just keywords.

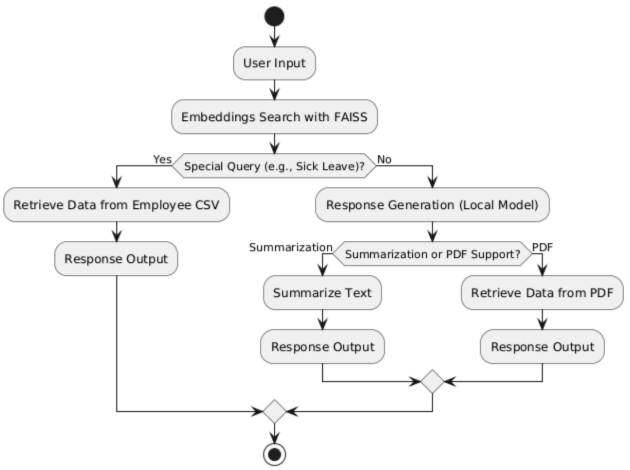
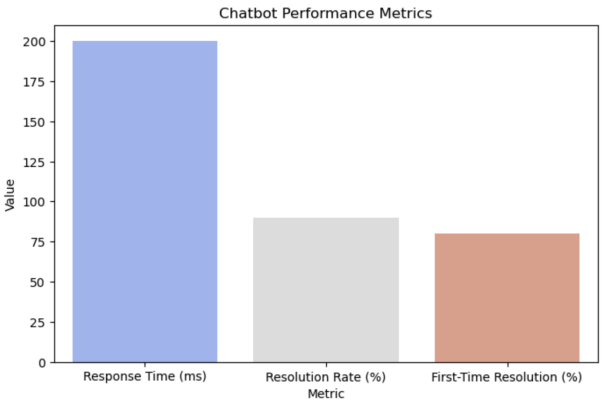


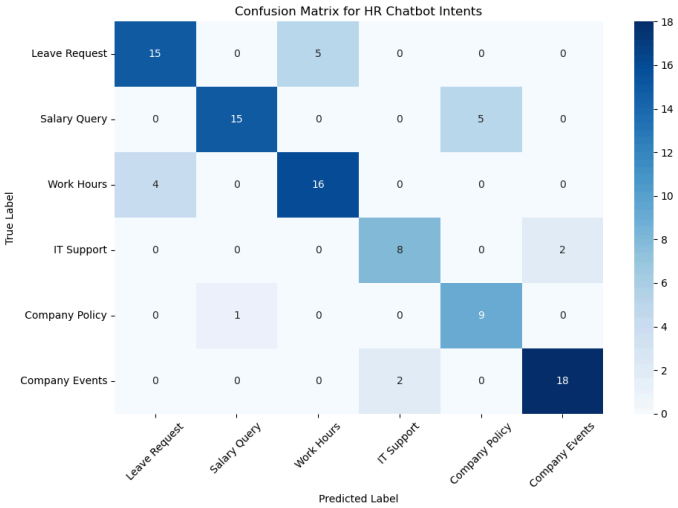
Figure 3: Flowchart of the system

- This flowchart outlines the process for generating responses based on user input, using an embeddings search powered by FAISS.
- After receiving User Input, the system performs an Embeddings Search with FAISS. Depending on the query type, the flow splits into two main branches.
- If the query is a Special Query (e.g., related to sick leave), the system retrieves relevant data from an Employee CSV and generates the Response Output.
- If the query does not involve a special case, the system proceeds to Response Generation using a local model.
- At this point, if the request involves summarization or querying a PDF, the system decides between two paths: Summarize Text (for text-based queries) or Retrieve Data from PDF (for PDF-related queries).
- Both paths result in the generation of the Response Output, after which the process concludes.
- The chart effectively shows how the system handles a variety of user queries, deciding whether to fetch employee data, summarize content, or extract PDF information to provide an appropriate response.

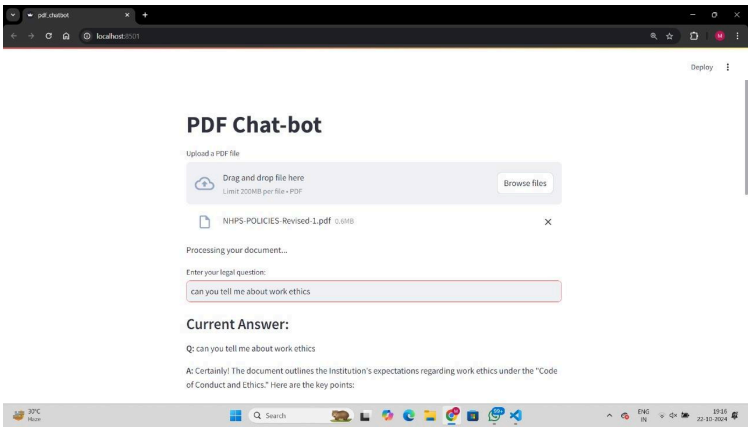
V. Results



This graph shows the Chatbot Performance Metrics like response time, resolution rate and first-time resolution.



This diagram shows the Confusion Matrix for HR Chatbot Intents. X axis shows the predicted label and the Y axis shows the true label.



This image shows the implementation of the PDF Chatbot.

VI. Conclusion

The SmartGuide Chatbot project successfully demonstrates the integration of advanced AI and natural language processing technologies to streamline employee support and document analysis within an organization. By combining features such as quick query handling, secure two-factor authentication (2FA), fast and scalable processing, and parallel operation, the chatbot offers an efficient and robust solution for HR and operational needs. Its ability to perform document processing, keyword extraction, and summarization further enhances the utility of the system by automating repetitive tasks and improving access to key information.

The SmartGuide Chatbot's design prioritizes user experience by ensuring quick response times, reliable data retrieval, and secure interactions. The project showcases how AI-driven solutions can simplify

workflows, improve operational efficiency, and offer scalable support for organizations handling large amounts of employee queries and documents. As such, this project marks an important step forward in building intelligent, automated systems capable of improving the overall productivity and efficiency of organizations.

VII. 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.yorksj.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.