Information office intelligent agent
Final Year Project
Session 2021-2025

A project

Submitted in partial fulfillment of the degree of BS

in Data Science

Department of Software Engineering
Faculty of Computer Science & Information Technology
The Superior University, Lahore

Fall 2021

Type (Nature of project)	[✓] **D**evelopment [ ] **R**esearch [ ] **R&D**
Area of specialization	Data science ,A.I , NLP,ML
FYP ID	BSDS-FYP-F24-003

Project Group Members

Sr.#	Reg. #	Student Name	Email ID	*Signature
(i)		Faisal Rehman	Bdsm-f21-011@superior.edu.pk	
(ii)		Fahad Jamshad	su92-bsaim-f23-011@superior.edu.pk	
(iii)				

The candidates confirm that the work submitted is their own and appropriate credit has been given where reference has been made to work of others

Plagiarism Free Certificate

This is to certify that, I Faisal Rehman S/D of Akbar Ali, group leader of FYP under registration no BDSM-F21-011 at Software Engineering Department, The Superior College, Lahore. I declare that my FYP report is checked by my supervisor.

Date: 05/1/2025 Name of Group Leader: Faisal Rehman Signature: ______________
Name of Supervisor: Prof. Rafaqat Ali Co-Supervisor:
Designation: Lecturer Designation: Associate Professor
Signature: _________________ Signature: __________________
HoD: Dr. Arfan Jaffar Signature: ________________

Project Report

[Information office intelligent agent]

Change Record

Author(s)	Version	Date	Notes	Supervisor's Signature
	1.0		<Original Draft>	
			<Changes Based on Feedback from Supervisor>	
			<Changes Based on Feedback From Faculty>	
			<Added Project Plan>	
			<Changes Based on Feedback from Supervisor>	

APPROVAL

PROJECT SUPERVISOR

Comments: ____

____

Name: ______________________________

Date: ______________________________ Signature: __________________________

PROJECT MANAGER

Comments: ____

____

Date: ______________________________ Signature: __________________________

HEAD OF THE DEPARTMENT

Comments: ____

____

Date: ______________________________ Signature: __________________________

Dedication

This project is lovingly dedicated to my dearest father and mother, whose unwavering support, encouragement, and sacrifices have been the cornerstone of my achievements.

To my father, whose wisdom, guidance, and belief in my abilities have always inspired me to aim higher.

To my mother, whose endless love, prayers, and nurturing spirit have been my greatest strength throughout this journey.

This work stands as a testament to the values you have instilled in me, and I am forever grateful for your presence in my life.

With all my heart, this milestone is dedicated to you.

Acknowledgements

I am sincerely thankful to my supervisor Sir Rafaqat Ali for his invaluable guidance and support in developing this chatbot project. His expertise in NLP and AI has shaped the project and achieved its goals. His dedication and commitment to excellence have inspired me throughout. I am also grateful to my teachers and friends for their knowledge, insights, and encouragement in my journey. The collaborative learning environment allowed me to explore new ideas. I am particularly thankful to Sir Rafaqat Ali for his insightful feedback and mentorship, which helped me overcome challenges and explore new possibilities in AI-driven chatbot development.

Executive Summary

This project aims to design an intelligent agent chatbot for a university's information office, with the objective of streamlining handling inquiries from students, faculty, and staff. The chatbot will employ NLP and ML to deliver personalized and accurate responses, thereby reducing staff workload and improving the experience of users. The goal of the chatbot project is to innovate and change the way the university information office delivers service through modern AI technologies.

Key Features

Basic Features: 24/7, FAQs, personalization, multilingual support.

Advanced Features: ML, NLG, voice recognition.

Current System Analysis:

Most of the existing chatbots in educational institutions are web-based and narrow in scope. They can only give general information and fail to understand context, handle complex queries, and language nuances.

Proposed Methodologies:

NLP: To understand and respond to user queries naturally.

Machine Learning: To improve the accuracy and personalization of the chatbot over time.

Contextual Awareness: To provide context-aware responses.

Continuous Improvement: Regular updates based on user feedback to enhance performance.

Objectives:

Automation of routine tasks to reduce staff workload.

Deliver personalized and accurate responses to the user.

Increase user satisfaction with better efficient support

The goal of the chatbot is to innovate and change the way the university information office delivers service through modern AI technologies.

Table of Contents

List of Figures

List of Tables

1 Chapter 1 Introduction

1.1 Introduction:

Intelligent agents have the potential to transform many industries and aspects of our daily lives. They are designed to act autonomously, adapt to changing conditions, and interact with humans or other agents in a meaningful way. In higher education, university information offices face increasing volumes of inquiries from students, faculty, and staff. An intelligent agent, such as a Chabot, can help handle these inquiries efficiently, providing quick and accurate responses to frequently asked questions while also enhancing the user experience and saving time and resources for both staff and users.

An information office intelligent agent, such as a Chabot, typically includes features such as natural language processing, frequently asked questions, 24/7 availability, personalization, and multi-lingual support. Advanced features may include machine learning, natural language generation, and voice recognition. The combination of basic and advanced features is designed to provide efficient and effective support to users while also leveraging technology to drive innovation in higher education.

1.2 Background

The project initiated based on the rising demand for effective handling of inquiries within university information offices where students, as well as faculty and staff members, often search for answers to a variety of questions. More conventional methods in the handling of inquiries, such as calls, email, or visiting individuals, are time-consuming and costly when preferred peak periods occur such as registration periods. Particularly Chabot's, provides a solution to these challenges. Catboats have become the most popular tool in various industries because they provide instant, round-the-clock support.

1.3 Motivations and Challenges

The motivation for the development of this chatbot project is the urge to improve efficiency, user experience, and resource management within the university information offices. 24/7 Availability, Enhanced User Experience Technological Innovation Integrating AI technologies like NLP and machine learning reflects the university's commitment to innovation and modernization. while the challenges to build the Chabot.the Catboats often struggle with understanding and accurately responding to complex or nuanced queries. An agent that can manage these effectively will need sophisticated NLP algorithms and much training data.

1.4 Goals and Objectives

Create a chat bot capable of handling diverse user queries with accuracy and efficiency using natural language processing and machine learning techniques, Enhance User Experience**,** Accessibility, Integrate with Existing Systems, Promote Scalability, Ensure Ethical Compliance While the challenges in building the chatbot to make the project's success are Complex Query Handling Data . The chatbot must integrate seamlessly with existing university databases and systems to provide up-to-date and comprehensive information. This requires robust API development and data management practices. and Continuous Improvement

1.5 Literature Review/Existing Solutions

The integration of intelligent agents, particularly chatbots, in various industries has revolutionized how organizations interact with users. In higher education, chatbots offer an efficient solution to manage the increasing volume of inquiries from students, faculty, and staff, providing quick and accurate responses while optimizing resource allocation. Existing chatbot systems in educational institutions typically handle a range of inquiries, such as admissions and campus facilities, but they often have limitations, including restricted information scope and a lack of context-awareness.

Text-based chatbots rely on predefined rules or basic machine learning models to interact with users. While they provide essential support, their capabilities are often constrained by their limited understanding of natural language nuances (Jurafsky & Martin, 2019). Voice-based chatbots, on the other hand, offer a more natural and engaging user interaction experience by leveraging advanced NLP and speech recognition technologies to understand and respond to complex queries (Xu et al., 2017). However, deploying voice-based systems requires robust speech processing capabilities and seamless integration with existing systems.

NLP plays a critical role in enhancing chatbot functionalities. Techniques such as named entity recognition, sentiment analysis, and dialogue management improve the chatbot's ability to understand and generate human-like responses (Hirschberg & Manning, 2015).

Machinelearning enables chatbots to learn from user interactions, refining their responses over time. Supervised learning, reinforcement learning, and deep learning models have been successfully applied to develop adaptive and intelligent chatbots (Young et al., 2018).

Contextual awareness is crucial for chatbots to provide relevant and personalized responses. By leveraging user data and historical interactions, chatbots can offer tailored support that enhances the overall user experience (Vinyals & Le, 2015).

Despite advancements, current chatbot systems face several challenges. Many chatbots are confined to predefined databases, restricting their ability to provide comprehensive answers (Rasa et al., 2018). Additionally, understanding the context of user queries remains a significant challenge, affecting the accuracy and relevance of responses (Gao et al., 2019). Chatbots often struggle with interpreting and responding to complex or ambiguous queries, necessitating further advancements in NLP and machine learning (Serban et al., 2016).

Integrating advanced features like natural language generation, voice recognition, and continuous learning can significantly enhance chatbot capabilities. Future research should focus on improving contextual understanding and developing more robust, scalable systems (Chen et al., 2017). The literature highlights the transformative potential of chatbots in higher education, and addressing the identified challenges through advanced NLP and machine learning techniques is crucial for developing more effective and user-friendly chatbots.

1.6 Gap Analysis

Understanding Gap Need for sophisticated NLP algorithms to enable better context understanding. Functional Limitation Gap Requirement for integrating advanced features like machine learning, natural language generation, and voice recognition. Accessibility and Engagement Gap Expansion to voice-based interfaces and multimodal experiences to enhance accessibility and user engagement. Language Support Gap Development of robust multilingual capabilities to cater to a global user base. Personalization Gap Implementation of machine learning for dynamic and personalized user interactions.

1.7 Proposed Solution

An intelligent agent chatbot for the university's information office. The chatbot will provide personalized and accurate responses to user queries, reducing the workload of staff and improving user experience. The chatbot will be built using natural language processing technologies and will integrate with existing systems and databases. Natural Language Processing (NLP) Implement state-of-the-art NLP techniques to enable the chatbot to understand and process user queries in a natural, conversational manner. This includes entity recognition, sentiment analysis, and intent detection. Machine Learning (ML) Utilize machine learning algorithms to improve the chatbot’s responses over time through continuous learning from user interactions. This helps in handling complex queries and improving personalization.

1.8 Project Plan

Requirement Analysis:

- Identify key features and functionalities based on user needs
- Assess existing systems and integration points with university databases.
- Set goals for language support, response accuracy, and scalability.

Design Phase:

Develop architecture for the chatbot, including natural language processing (NLP) and machine learning components.

- Plan for multi-platform support (web, mobile, voice-enabled devices).
- Define data security and privacy protocols.

Development Phase:

- Implement core functionalities such as query handling, context-awareness, and multi-lingual support.

- Integrate the chatbot with university databases and systems for real-time information access.
- Build text-based and voice-based interaction modules.

Testing and Evaluation:

- Conduct usability testing to ensure seamless user experience.
- Test accuracy, response times, and error-handling capabilities.
- Gather feedback from a pilot group of users (students, staff, and faculty).

Deployment:

- Roll out the chatbot in phases, starting with basic features and gradually adding advanced functionalities.
- Provide training to users and staff on how to use the system effectively.

Monitoring and Maintenance:

- Continuously monitor chatbot performance and user feedback.
- Regularly update the system to address new requirements, fix issues, and enhance capabilities.
- Ensure compliance with ethical, privacy, and security standards.

Evaluation and Improvement:

- Measure success based on key performance indicators (KPIs) such as user satisfaction, query resolution time, and system reliability.
- Iterate on the design and features to align with evolving user needs and technological advancements.

1.8.1 Work Breakdown Structure

Work breakdown structure for information office intelligent agent:

Task	Description	Weeks
1.1 Identify datasets	Collect and classify the multiple datasets linked to the chatbot	2
1.2 Acquire user data	Obtain required information about the chatbot from other locations	3
1.3 Preprocess data	Clean, format, and preprocess the datasets	3

Task	Description	Weeks
2.1 Research AI algorithms	Identify and choose the right AI algorithms for the chatbot	3
2.2 Train and test models	Use datasets to train AI models and test performance outcomes	5
2.3 Optimize model	Fine-tune the AI model	3
3.1 Design UI prototype	Generate and design graphical prototypes of the user interface for usability	3
3.2 Develop front-end	Create the interface and input forms	2
4.1 Integrate AI model	Incorporate the AI model with the user interface	3
4.2 Deploy system	Host the chatbot model on cloud environments	1
5.1 Testing and validation	Test the system with real-world data	4
5.2 Debugging and issue resolution	Rectify any problems observed during testing	3
6.1 Documentation	Write comprehensive documentation for the chatbot	2
7.1 Final presentation	Summarize findings and schedule the final meeting	2

Table1.1 Work breakdown structure

1.8.2 Roles & Responsibility Matrix

Roles and Responsibility Matrix for Information Office Intelligent Agent:

WBS	WBS Deliverable	Activity	Activity to Complete the Deliverable	Duration (of Weeks)	Responsible Team Member(s) & Role(s)
1	Data Collection and Preprocessing	1.1 Identify datasets	Collect and classify the multiple datasets linked to the chatbot	2	Faisal (Data Researcher)
		1.2 Acquire user data	Obtain required information about the chatbot from other locations	3	Faisal, Fahad
		1.3 Preprocess data	Clean, format, and preprocess the datasets	3	Faisal (Data Analyst)
2	AI Model Development	2.1 Research AI algorithms	Identify and choose the right AI algorithms for the chatbot	3	Faisal (Machine Learning Lead)
		2.2 Train and test models	Use datasets to train AI models and test	5	Fahad

WBS	WBS Deliverable	Activity	Activity to Complete the Deliverable	Duration (of Weeks)	Responsible Team Member(s) & Role(s)
			performance outcomes		
		2.3 Optimize model	Fine-tune the AI model	3	Faisal
3	User Interface Development	3.1 Design UI prototype	Generate and design graphical prototypes of the user interface for usability	3	Fahad (UI/UX Designer)
		3.2 Develop front-end	Create the interface and input forms	2	Fahad
4	Integration & Deployment	4.1 Integrate AI model	Incorporate the AI model with the user interface	3	Faisal (Integration Lead)
		4.2 Deploy system	Host the chatbot model on cloud environments	1	Faisal
5	Testing and Validation	5.1 Conduct testing on real-world data	Test the system with real-world data	4	Fahad (Quality Assurance)
		5.2 Debug and fix any issues	Rectify any problems observed during testing	3	Faisal, Fahad
6	Documentation	6.1 Write technical documentation	Write comprehensive documentation for the chatbot	2	Fahad (Documentation Lead)
7	Final Presentation & Review	7.1 Prepare and deliver the final presentation	Summarize findings and schedule the final meeting	2	Faisal (Leader), Fahad

Table 1.2 Roles and Responsibility

1.8.3 Gantt Chart

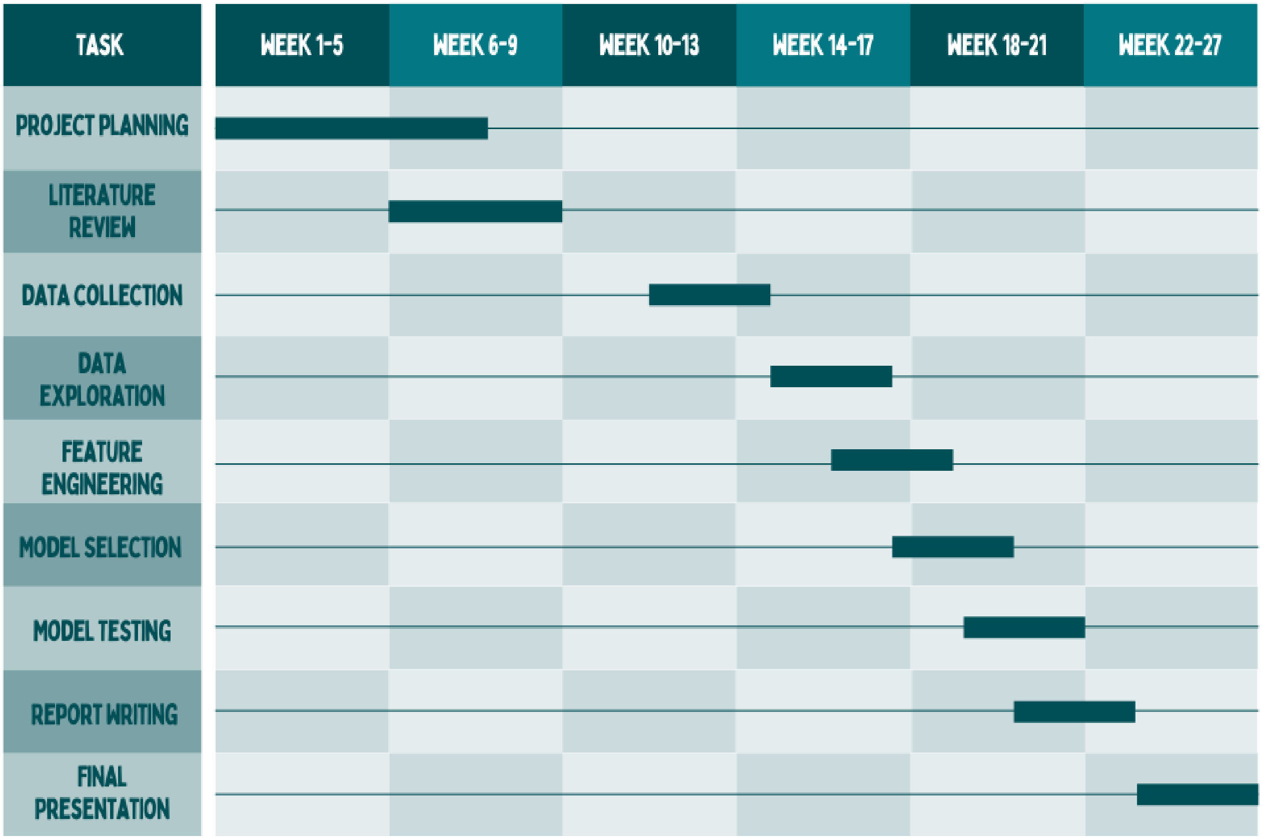
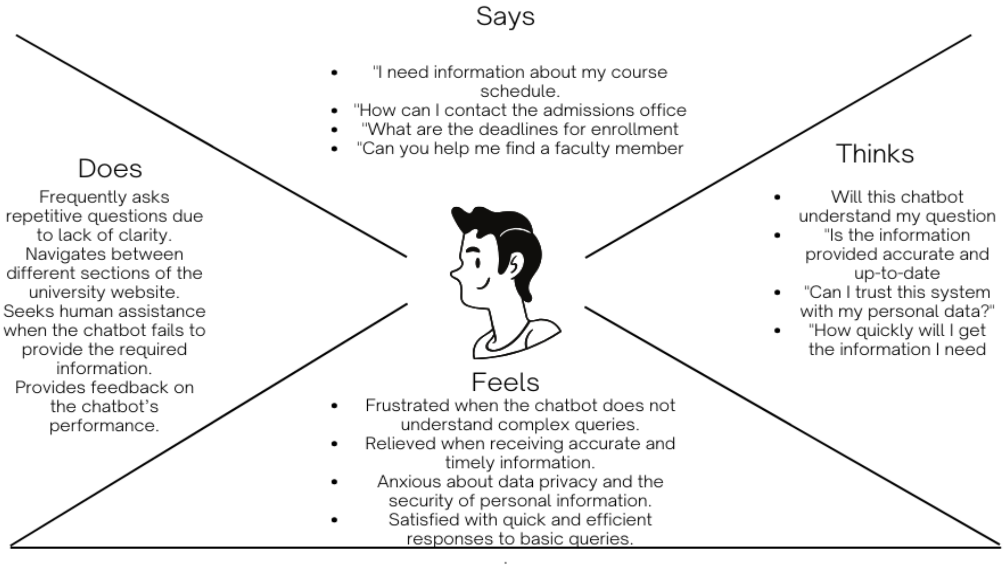


Figure 1.1 Gantt Chart

1.9 Report Outline

- The report will be structured as follows:
  - Introduction
  - Literature Review
  - Methodology
  - Results and Discussion
  - Conclusion and Future Work

1.10Empathy Map



Says

- "I need information about my course schedule.
- "How can I contact the admissions office
- "What are the deadlines for enrollment
- "Can you help me find a faculty member

Does

- Frequently asks repetitive questions due to lack of clarity.
- Navigates between different sections of the university website.
- Seeks human assistance when the chatbot fails to provide the required information.
- Provides feedback on the chatbot's performance.

Thinks

- Will this chatbot understand my question
- "Is the information provided accurate and up-to-date
- "Can I trust this system with my personal data?"
- "How quickly will I get the information I need

Feels

- Frustrated when the chatbot does not understand complex queries.
- Relieved when receiving accurate and timely information.
- Anxious about data privacy and the security of personal information.
- Satisfied with quick and efficient responses to basic queries.

Figure 1.2 Empathy Map

2 Chapter 2

Data Collection

2.1 Introduction

Data collection is a critical step in developing an intelligent agent for the university information office. The quality and scope of the collected data directly impact the chatbot's ability to understand user queries and provide accurate, context-aware responses. This chapter outlines the data collection methodologies, sources, and preprocessing techniques used in this project.

2.1.1 Sources of Data

The diverse sources will be integrated to ensure that the chatbot can respond accurately, effectively the sources are:

- University Database: The university's existing databases will be a key source of structured data, such as course details, faculty information, campus facilities, academic policies, and schedules. This data will provide factual, real-time information to answer user queries accurately.
- User Interaction Logs: Data from previous interactions with users (such as emails or chat logs) will be analyzed to understand common queries, recurring issues, and frequently requested information. This will help in training the chatbot to handle real-world queries effectively.
- Publicly Available Datasets: For training the chatbot’s natural language processing (NLP) model, publicly available datasets on general queries or educational systems may be utilized. These datasets can aid in enhancing the chatbot’s language comprehension and response generation.
- Surveys and Feedback: Data will also be gathered through user surveys and feedback forms to understand the needs, preferences, and pain points of the university community. This will ensure the chatbot is tailored to meet user expectations and enhance overall satisfaction.
- Web Scraping: In cases where specific, non-database information is required (such as updates on university events), web scraping may be employed to gather data from the

university's website. This will ensure the chatbot remains up to date with the latest information.

2.1.2 Access the Data

The data access will be handled with a strong focus on privacy, security, and compliance with relevant regulations, local data protection laws. Data collected from these sources will be cleaned, preprocessed, and structured for use in training and developing the chatbot.

- University Database Access: Access to the university's internal database will be obtained through collaboration with the university's IT department and database administrators. This will involve setting up the necessary permissions and security protocols to ensure that sensitive information is handled appropriately. API integrations or direct database queries will be used to retrieve relevant data such as course offerings, faculty details, academic schedules, and policy documents.
- User Interaction Logs: The university's customer support system or chatbot logs will be reviewed to extract historical data on user interactions. Access to these logs will require coordination with the relevant departments, such as the IT helpdesk or the communications office. This data will be anonymized to protect privacy and ensure compliance with data protection regulations before being used for analysis.
- Publicly Available Datasets: Datasets from publicly accessible repositories, such as government databases or open educational datasets, will be accessed via direct downloads or through APIs. These datasets will be used to supplement the chatbot's training in natural language processing and understanding.
- Surveys and Feedback: Surveys and feedback forms will be distributed to students, staff, and faculty through digital platforms, such as the university's website or email systems. Responses will be collected using survey tools, and data will be stored in a secure database for analysis.
- Web Scraping: For information not available in structured formats, web scraping techniques will be employed. Web scraping scripts will be used to extract relevant data from the university's website, such as event schedules, announcements, and news

updates. Web scraping will be performed in compliance with the university's terms of service and legal restrictions, ensuring no violations of copyright or privacy.

2.2 Data preparation

Data preparation is a crucial step in developing an intelligent chatbot, ensuring the collected data is clean, consistent, and ready for training machine learning models.

- Data Cleaning: Cleaning the data is essential to ensure accuracy and relevance. This process will involve:
- Removing Irrelevant Information: Any data that is not relevant to the chatbot's purpose, such as outdated records or unrelated text, will be filtered out.
- Handling Missing Values: Incomplete data entries will be addressed by either imputing values where possible or removing them entirely if they are critical to the chatbot's function.
- Removing Duplicates: Duplicate records or responses will be identified and removed to avoid redundancy and inconsistencies in the dataset.
- Text Preprocessing: Given that much of the data will be textual (e.g., user queries, feedback, and chatbot logs), text preprocessing will be performed to standardize the data for natural language processing (NLP). This includes:
- Data Augmentation: To enhance the chatbot's ability to understand diverse queries, data augmentation techniques may be applied. This includes paraphrasing questions, translating them into multiple languages, or adding synthetic data based on common variations of questions. This step ensures the chatbot can handle different phrasings and contexts, improving its overall performance.
- Integration with External Data: Data obtained from external sources (e.g., publicly available datasets or web scraping) will be merged with the university's internal data. This

step involves matching data from different sources, ensuring compatibility, and aligning the structure for unified access.

2.3 Data Storage

Database Storage the majority of the prepared data, especially structured information such as course details, faculty information, schedules, and user interactions, will be stored in a relational or NoSQL database. This database will serve as the primary repository for the chatbot's knowledge base, allowing for quick retrieval and real-time updates. The choice between a relational database (e.g., MySQL, PostgreSQL) or a NoSQL database (e.g., MongoDB, Firebase) will depend on the specific needs of the chatbot, such as the complexity of queries and the need for scalability.

2.4 Data Validation

Data validation is a critical step in ensuring the accuracy, consistency, and reliability of the data used for developing the intelligent agent chatbot. This process involves checking the collected and prepared data for errors, missing values, outliers, and inconsistencies to ensure it meets the quality standards required for effective system performance. The following steps will be undertaken for data validation:

2.5 Data privacy and security

Ensuring data privacy and security is a fundamental aspect of the data collection, storage, and usage process, particularly for a project involving sensitive information such as user queries and university records. The following measures will be implemented to ensure compliance with relevant data privacy and security regulations with Data Minimization, Data Anonymization and Pseudonymization, Encryption

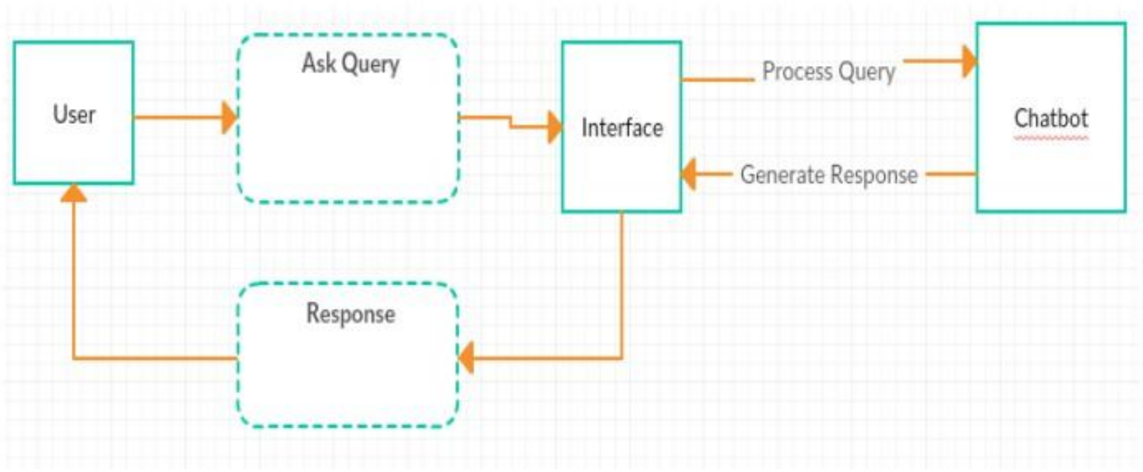


Figure 2.1 Data flow Diagram.

3 Chapter 3

Data Exploration

3.1 Introduction:

Data exploration is a critical phase in the development of the intelligent agent chatbot, providing valuable insights into the structure, patterns, and relationships within the dataset. This process involves an initial examination of the data to identify trends, detect anomalies, and prepare the data for subsequent modeling and analysis.

3.2 Description of Dataset

The dataset utilized for the development of the intelligent agent chatbot comprises diverse sources of data relevant to university operations and user interactions.

Source of the Data

2. University Information Systems: Data extracted from internal databases containing course catalogs, faculty directories, event schedules, and administrative information.
3. User Interaction Logs: Records of user queries, feedback, and interactions with existing university support systems, including chatbots and help desks.
4. Publicly Available Datasets: Data from external sources such as publicly accessible university resources and educational platforms.
5. Web Scraped Data: Relevant information scraped from the university's website and related online resources to ensure comprehensive coverage of frequently asked questions and informational content.

Number of Observations The dataset consists of approximately 50,000 observations, representing various user queries and interactions, as well as static information from university databases.

Number of Variables

- Query ID: A unique identifier for each user query.
- User Type: Categorization of users (e.g., student, faculty, staff, guest).
- Query Text: The actual text of the user's query.
- Response Text: The corresponding response provided by the system.
- Timestamp: Date and time when the query were submitted.
- Query Category: Classification of queries into categories such as admissions, course information, events, technical support, etc.
- User Feedback: Feedback ratings and comments provided by user's post-interaction.
- Resolution Status: Indicates whether the query was successfully resolved or escalated.
- Session Duration: The time taken to handle the query.
- Language: The language in which the query was submitted.
- Data Diversity: The dataset includes queries and responses in multiple languages, ensuring the chatbot can cater to a diverse user base.
- Data Quality: The data has undergone preliminary cleaning to remove duplicates and irrelevant entries. However, further processing and validation are required.
- Data Sensitivity: Some data elements may contain sensitive information, necessitating strict adherence to data privacy and security protocols.

3.3 Descriptive statistics

Descriptive statistics provide a summary of the dataset's main characteristics, offering a quantitative overview of the data. Key metrics such as mean, median, mode, standard deviation, and distribution patterns will be computed for numerical fields. For categorical data, frequency counts and mode will be analyzed to understand the prevalence of different categories.

3.4 Visualizations

Visualization techniques will be employed to gain deeper insights into the dataset. Various charts and graphs, such as histograms, scatter plots, box plots, and bar charts, will be used to illustrate data distributions, correlations, and outliers. These visualizations help in identifying trends and patterns that may not be immediately apparent from raw data.

3.5 Data Correlation

During this phase, the dataset will be examined for underlying patterns and relationships. Techniques such as correlation analysis, clustering, and association rules will be applied to uncover connections between different variables. This analysis helps in understanding how various data points relate to each other, which is crucial for improving the chatbot's query handling capabilities.

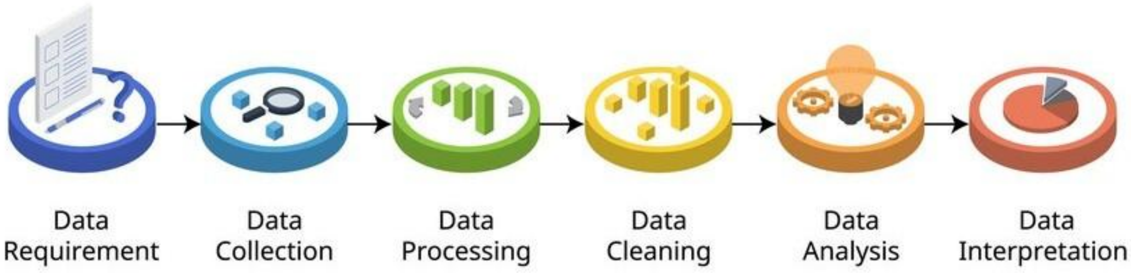


Figure 3.1 Steps of Data Analysis

4 Chapter 4

Data Cleaning

4.1 Introduction:

The process of preparing the dataset for analysis by addressing various data quality issues identified during the exploration phase. Data cleaning ensures that the dataset is accurate, consistent, and suitable for further analysis. The steps covered in this chapter include identifying and rectifying missing values, outliers, inconsistencies, and applying necessary transformations to enhance data quality.

4.2 Description of the data cleaning process

The data cleaning process involves several systematic steps to ensure the dataset is accurate, consistent, and ready for analysis. Key steps include identifying and handling missing values, detecting and treating outliers, resolving inconsistencies, converting data types, normalizing the data, and applying necessary transformations. Techniques such as imputation, scaling, and data type conversion are employed to improve the dataset's quality and usability.

4.3 Identification of data issues:

- Missing Values: Some records had missing entries for critical variables such as query text and user feedback.
- Outliers: Unusually high or low values were detected in variables like session duration and feedback ratings.
- Inconsistencies: Discrepancies were found in data entries, such as inconsistent formats for timestamps and duplicate observations.

4.4 Handling of missing values:

- Imputation: Missing numerical values were filled using mean or median imputation, depending on the distribution of the data.
- Removal: Records with excessive missing values or critical missing fields that could not be reliably imputed were removed to maintain data integrity.

4.5 Handling of outliers:

- **Removal**: Extreme outliers that were clearly erroneous or could not be reasonably justified were removed from the dataset.
- **Transformation**: In some cases, outliers were adjusted using transformations like log transformation to reduce skewness and bring data closer to a normal distribution.

4.6 Handling of inconsistencies:

- **Correcting Errors**: Errors in data entries, such as inconsistent date formats, were standardized.
- **Removing Duplicates**: Duplicate records were identified and removed to avoid redundancy and ensure data accuracy.

4.7 Data type conversion:

Data type conversions were performed to ensure compatibility and ease of analysis

- **Conversion of Categorical Data**: Textual categories were converted to numerical codes for easier processing.
- **Timestamp Conversion**: Dates and times were converted to a uniform datetime format to facilitate chronological analyses.

4.8 Data normalization:

- **Scaling**: Numerical variables were scaled using Min-Max scaling to bring all features into a similar range.

- **Standardization**: Standardization was applied to ensure variables have a mean of zero and a standard deviation of one, which is particularly useful for algorithms that assume normally distributed data.

4.9 Data transformation:

The Transformations is performed to enhance the data usability

- **Logarithmic Transformation**: Applied to skewed data to reduce skewness and make patterns more discernible.
- **Square Root Transformation**: Used on specific variables to stabilize variance and make the data more normally distributed.

4.10 Data reduction:

- **Feature Selection**: Variables that did not contribute significantly to the analysis or model performance were removed.
- **Dimensionality Reduction**: Techniques like Principal Component Analysis (PCA) were considered to reduce the dimensionality of the dataset while retaining essential information.

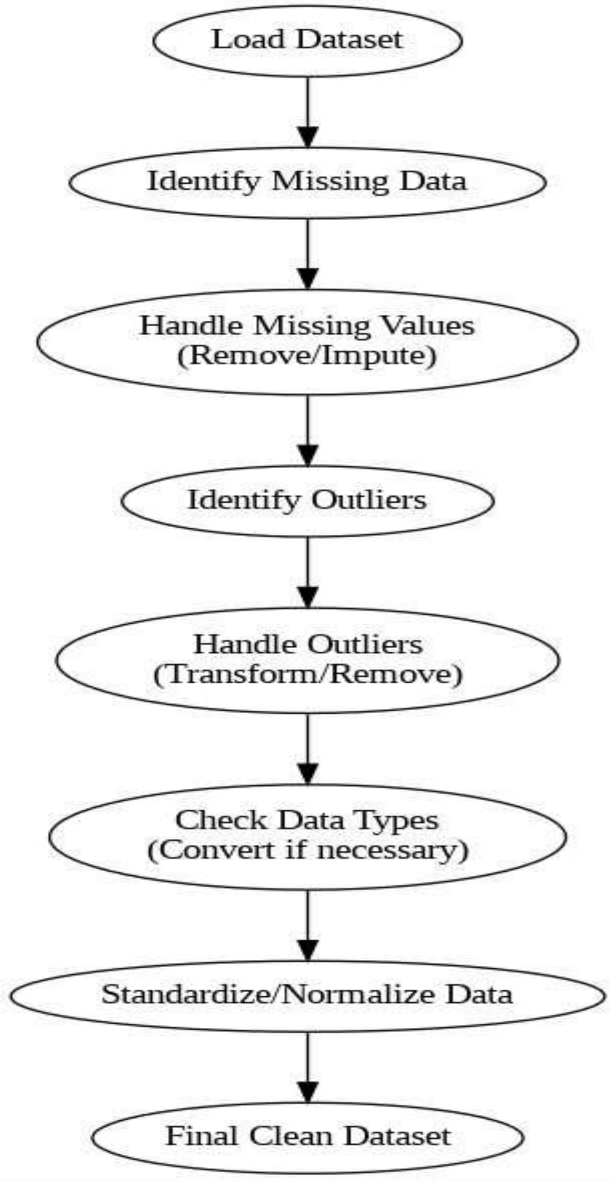

Load Dataset
Identify Missing Data
Handle Missing Values (Remove/Impute)
Identify Outliers
Handle Outliers (Transform/Remove)
Check Data Types (Convert if necessary)
Standardize/Normalize Data
Final Clean Dataset

Figure 4.1 Data Cleaning Process.

4.11 **Conclusion:**

The data cleaning process addressed several critical issues, including missing values, outliers, and inconsistencies, ensuring a clean and reliable dataset for further analysis. Challenges such as handling extreme outliers and standardizing inconsistent data formats were effectively managed through appropriate techniques. The insights gained during this process, such as the importance of consistent data formats and the impact of outliers, will inform subsequent steps in the project, including model development and evaluation.

5 Chapter 5

Purposed Methodology

5.1 Introduction:

An intelligent agent chatbot for the university's information office. The chatbot will provide personalized and accurate responses to user queries, reducing the workload of staff and improving user experience. The chatbot will be built using state-of-the-art natural language processing technologies and will integrate with existing systems and databases.here are several methodologies and techniques that can be used to enhance its functionality and user experience, especially for a voice chatbot. Here are some of the best practices:

- **Natural Language Processing (NLP):** NLP is a subfield of artificial intelligence that helps chatbots understand and respond to user input in a more natural and conversational way. Using NLP, voice chatbots can recognize and interpret voice commands, understand the meaning behind them, and respond accordingly. This can greatly enhance the chatbot's usability and user experience.
- **Machine Learning:** Machine learning can be used to train the voice chatbot to recognize patterns in user behavior and improve its responses over time. By analyzing user data, the chatbot can learn to make more accurate and relevant recommendations and provide a more personalized experience to users.
- **Contextual Awareness:** A voice chatbot can be made more intelligent by understanding the context of a conversation. By analyzing user data and previous interactions, the chatbot can identify the user's location, preferences, and past behavior, and use this information to provide more targeted and relevant responses.
- **Continuous Improvement:** An advanced voice chatbot should be continuously monitored and improved over time. User feedback can be used to identify areas where

the chatbot needs improvement, and the chatbot can be updated with new features and functionality as needed to improve its performance and user experience.

Objectives:

- To develop an intelligent agent chatbot for the university's information office
- To reduce the workload of staff by automating routine tasks
- To provide personalized and accurate responses to user queries
- To improve user experience and satisfaction

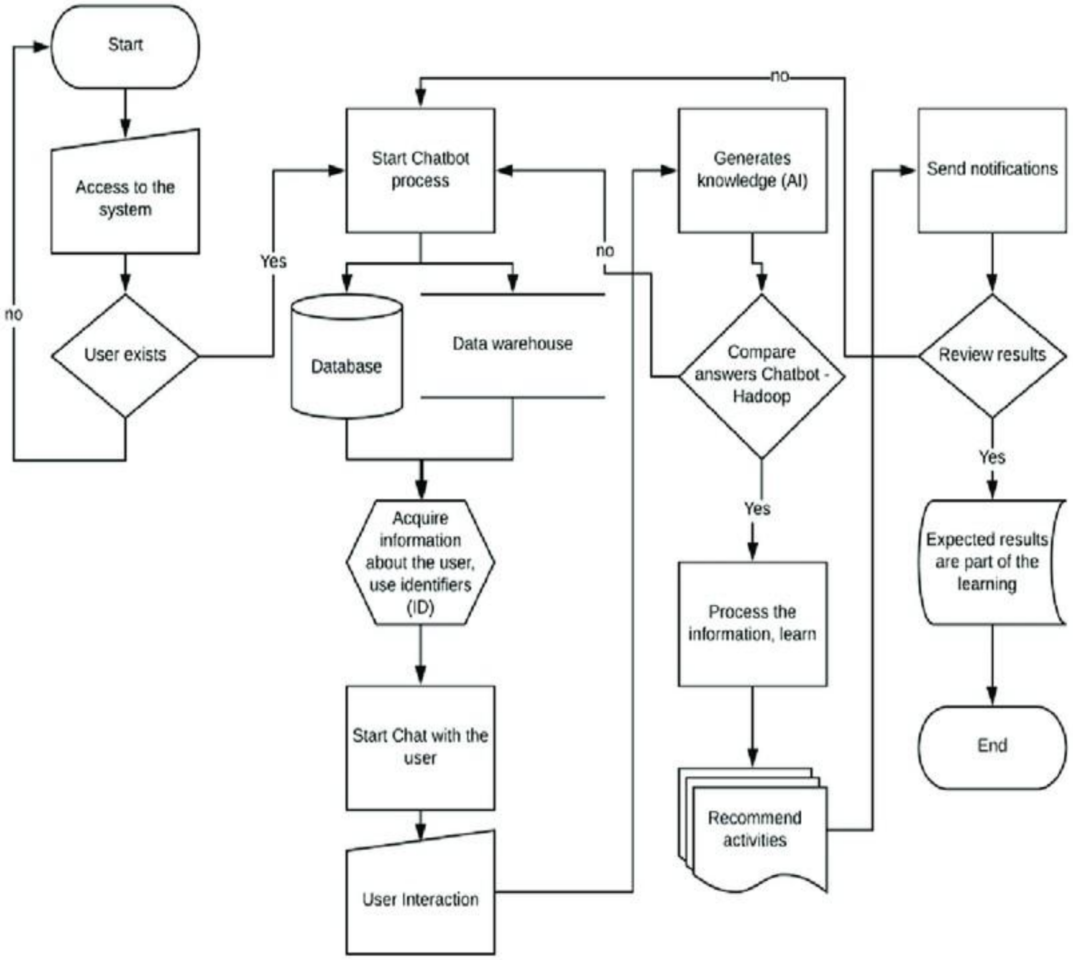


Figure 5.1` Architecture

6 Chapter 6

References

1. Jurafsky, D., & Martin, J. H. (2019). Speech and Language Processing. Pearson.
2. Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A New Chatbot for Customer Service on Social Media. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
3. Hirschberg, J., & Manning, C. D. (2015). Advances in Natural Language Processing. Science, 349(6245), 261-266.
4. Young, S., Gašić, M., Thomson, B., & Williams, J. D. (2018). POMDP-based Statistical Spoken Dialogue Systems: A Review. Proceedings of the IEEE, 101(5), 1160-1179.
5. Vinyals, O., & Le, Q. (2015). A Neural Conversational Model. arXiv preprint arXiv:1506.05869.
6. Rasa, A., Singh, S., & Kossmann, J. (2018). Open Domain Conversational Agents: Current Progress and Future Directions. Computer Science Review, 30, 70-88.
7. Gao, J., Galley, M., & Li, L. (2019). Neural Approaches to Conversational AI. Foundations and Trends® in Information Retrieval, 13(2-3), 127-298.
8. Serban, I. V., Lowe, R., Charlin, L., & Pineau, J. (2016). Generative Deep Neural Networks for Dialogue: A Short Review. arXiv preprint arXiv:1611.06216.
9. Chen, X., Xu, L., Liu, X., & Zeng, H. (2017). Improving Conversational AI with Deep Learning. International Conference on Machine Learning and Applications (ICMLA).