INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

# Multilingual Dialogue Summary Generation System for Customer Services

Project Specification Design and Prototype

**Mr Tharindu De Silva**

Supervised by

**Ms Dileeka Alwis**

Submitted in partial fulfilment of the requirements for the BEng (Hons) Software Engineering degree at the University of Westminster.

**February 2023**

# Abstract

A written or spoken conversation between two or more people is known as a dialogue. In most of modern-day applications where the conversation happens, it generates lots of data either in textual format or audio format. Summarizing these data or dialogues is known as the dialogue summarization where it only focusses on the relevant information. Dialogue summarization aims summarize only the necessary information where a reader can quickly capture the highlights of the dialogue without reviewing the whole.

Customer service is a domain where it generates lot of data daily. Conversation between a support agent and the customer can take up to a lengthy dialogue. At the end of the dialogue conversation, support agents should write a short description (a summary) of the dialogue which will help for later reference. This is a time consuming and requires human resources.

MultiDialogSum is a dialogue summary generation system for customer services where it supports generating summaries for multiple languages. The use of recent cross-lingual transfer modals and machine translations techniques are utilized to achieve these capabilities. This will be developed and release as a web application for the users.

**Keywords**: Dialogue Summarization, Natural language processing, Cross-lingual transfer modals, Machine translation, Machine Learning

**Project Descriptors:**

Artificial Intelligence → Natural Language Processing → Language models, Machine translation

# Contents

## List of Figures

# List of Tables

# Chapter 1: Introduction

## 1.1 Chapter Overview

Introduction chapter will include the summary of this research project which is a dialogue summary generation system that support more languages. This chapter will include the information about the problem domain, research gap, research challenges and objectives that the author wishes to achieve by the end of project completion.

## 1.2 Problem Domain

A written or spoken conversation between two or more people is called a dialogue. Dialogue summarization is the technique of condensing a dialogue so that a reader can quickly understand the exchange. The dialogue summarization method involves extracting important information from the discourse to produce a summary highlighting the conversation's main points. With the development of communication technology in recent years, dialogues have become an important way of information exchange.

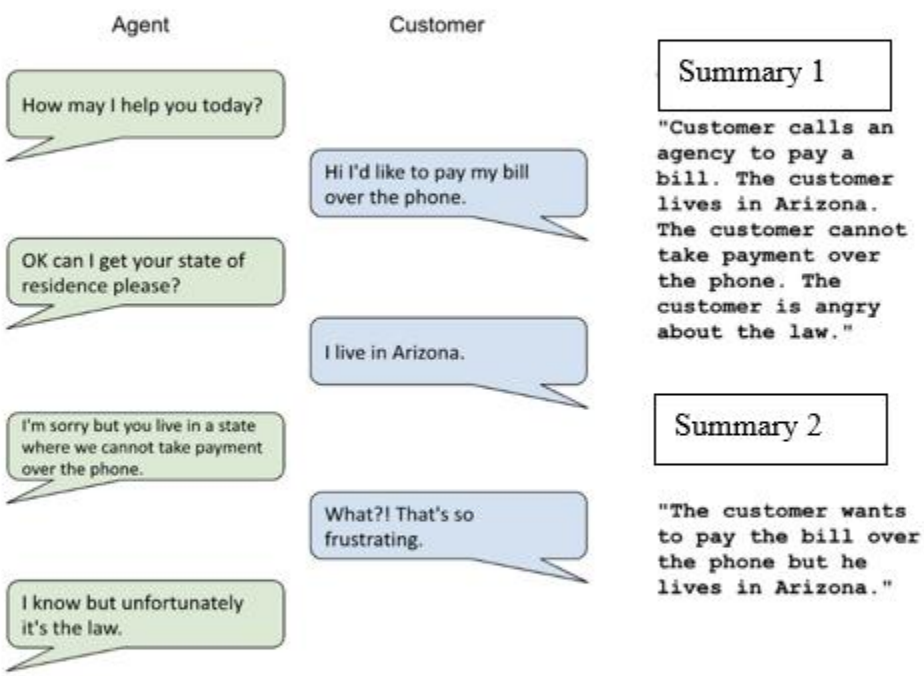



*Figure 1 - Dialogue Summary*

This proposal aims to provide the reader with an outline of dialogue summarization and how it has grown in recent years in different problem domains. The related work is discussed, which will identify and justify the existing research gap and challenges.

### 1.2.1 Impact of Dialogues on Communication Technology

Dialogues have been utilized within different domains as a communication technology for real-life applications. For example, some are in meetings, chats, interviews, and customer service. A dialogue can be represented in both textual and audible forms. Both are widely used in the real world. For example, chatbots are considered an intelligent way of communicating with minimal human interactions. These chatbots heavily use dialogues in a textual form as their primary source of communication. Another example is dialogues in the audible form in interviews. An interview between the two parties can be recorded and represented as a dialogue in an audible format.

### 1.2.2 The necessity of Dialogues Summarization

As previously mentioned, with these real-life applications' rapid growth, the dialogues' Summarization came into place because of various factors.

#### 1.2.2.1 Lengthy Dialogues

Dialogues can be used as a way of storing information for later reference. But the vast amount of data generation and length of the dialogues practically caused problems. In the customer service domain, a conversation between a customer and a support agent can be as follows. A customer can raise a complaint, and a support agent tries to solve the issue. End of the conversation, support agents are asked to write a summary of the problem and the proposed solution. So, this summary can be used by other agents who may have to deal with similar or the same customer issues without going through the entire dialogue.

#### 1.2.2.2 Time Consuming and Resource Cost

When it comes to massive data generation, analyzing a longer conversation requires human interaction and can cost much time and resources. Sometimes it is overwhelming to go through lengthy dialogues to understand what happened within the conversation.

## 1.3 Problem Definition

To automate dialogue summarization without human interaction, many researchers have put their effort into building solutions using machine learning and natural language processing technologies for this challenging problem because of their unique application value (Gao and Wan, 2022). However, these solutions are heavily focused on the high resource languages such as English, Chinese etc.(Feng, Feng and Qin, 2022). High-resources languages are known as languages in which many data resources exist.

With globalization's acceleration, a domain like customer service can be involved with multinational participants. Yet the current solutions are only capable of primarily handling the dialogues which are in English languages. This can be a critical situation where global customer service is required to summarize the dialogues between an agent and a customer using human resources with linguistic fluency in multiple languages.

### 1.3.1 Problem Statement

Dialogue summarization in customer service is not developed to handle languages other than mostly English. Customer services can always be engaged with multiple languages, yet there is no solution for multilingual dialogue summarization.

## 1.4 Aims and Objectives

### 1.4.1 Aims

*The aim of this research is to design, develop and evaluate a multilingual dialogue summary generation system for customer services using low linguistic resources with the help of the cross-lingual transfer method.*

To explain the research aim, this research project will focus on developing a system which is capable of utilizing low linguistic resources to build a multilingual dialogue summarizer. The recent development in pre-trained language models and the capabilities of cross-lingual transfer learning will be applied to achieve this. Cross-lingual transfer learning is the mechanism of learning and transferring knowledge from one natural language to another. The number of supporting languages of the multilingual dialogue summarizer can depend on the selected pre-trained model and its capabilities.

To prove or disprove the selected hypothesis, the necessary knowledge will be examined and researched through the project timeline, components will be developed, and the performance will be evaluated. This multilingual dialogue summary generation platform will be released on a hosted server for public use.

### 1.4.2 Research Objectives

| Research Objectives | Description | Learning Outcomes | Research Questions |
|---|---|---|---|
| Literature Review | Gather required material on previous work and critically evaluate the findings.<br>• **RO1:** Study on existing dialogue summarization techniques in the customer service domain.<br>• **RO2:** Study on existing methods of text summarization techniques.<br>• **RO3:** Conduct a preliminary study on pre-trained language models.<br>• **RO4:** Analyze the recent advancements in multilingual models. | LO1, LO4, LO8 | **RQ1, RQ2** |

| | | | |
|---|---|---|---|
| Requirement Elicitation | Determine the project's requirements using the proper methods to give a solution for the research problems and gaps should be handled based on relevant prior research knowledge.<br>• **RO1:** Gather information about requirements and resources related to dialogue summarization.<br>• **RO2:** Gather requirements related to the pre-trained language model and understand the capabilities of recent advancements.<br>• **RO3:** Gather information related to cross-lingual transfer learning. | LO2, LO6, LO8 | **RQ2, RQ3** |
| Design | Designing a system capable of generating a dialogue summary with multiple languages involved.<br>• **RO1**: To design a cross-lingual transfer learning model which can be implemented from the existing resources.<br>• **RO2:** To identify a suitable algorithm for the proposed methodology and design to summarize the dialogues. | LO1, LO3, LO5, LO8 | **RQ2, RQ3** |
| Implementation | Implementing a multilingual dialogue summarization platform.<br>• **RO1:** To train the cross-lingual model using the existing resources.<br>• **RO2:** Develop a model to summarize the dialogues. | LO1, LO5, LO7, LO8 | **RQ2** |

| | | | |
|---|---|---|---|
| | • **RO3:** Develop a user interface and which a user can interact with the application from a browser. | | |
| Evaluation | Testing the implemented system and dialogue summary generation with evaluation metrics.<br>• **RO1:** Test each component by creating unit tests.<br>• **RO2:** Using performance metrics to evaluate the effectiveness of the summary generation.<br>• **RO3:** To verify the functional and non-functional requirements were made after the research is completed. | LO1,<br>LO5, LO8 | **RQ2** |

*Table 1 - Research Objectives*

## 1.5 Novelty of the Research

### 1.5.1 Problem Novelty

Currently existing solution for dialogue summarization limited to a one language, but taking consideration that customer service is a globalized service and there is an urgent need to provide the dialogue summarization to a preferred language. (Feng, Feng and Qin, 2022a)

### 1.5.2 Solution Novelty

Developing a multilingual dialogue summarizer with use of cross-lingual transformer modal will be the proposed solution. During the past years cross-lingual transfer modal have been significantly improved for various natural language related works. With these recent advancements authors aim is to develop a multilingual dialogue summarizer which will provide a solution for the existing problem.

## 1.6 Research Gap

Based on the previous work on dialogue summarization, researchers have focused on the problem of low resources, which can be further divided into low linguistic resources and domain-specific data sets. Yet the approaches are heavily focused on English languages. This is due to the scarcity of the datasets, and the investment in such datasets can be costly (Zou et al., 2021). This can be addressed as a theoretical gap in the Dialogue summarization domain.

Dialogue summarization in customer service is where multiple languages get involved. But the currently available solutions can only perform an English dialogue summarization. This can be identified as an empirical gap in the customer service domain. This project focuses on the empirical gap in customer service and the theoretical gap in the Dialogue summarization domain.

## 1.7 Contribution to the Body of the Knowledge

Upon the completion of this research project, contributions can be summarized as follow:

- A platform to generate dialogue summary for customer service: Data Science [Machine Learning]
- A novel approach to use existing datasets for low resource languages: [Cross-Lingual Transfer, Pre-trained Language Models]

### 1.7.1 Contribution to Problem Domain

A platform that supports dialogue summary generation for multiple languages for the customer service domain will be explored. Considering current solutions' availability, a platform capable of generating summaries for more languages can be valuable for the growth of the customer service domain globally. The proposed solution can save both time and cost when it comes to preparing summarize for large quantity of dialogues without human resources.

### 1.7.2 Contribution to the Research Domain

A novel approach to utilize the existing datasets for dialogue summary generation and overcome the multilingual and scarcity of the datasets in low-resource languages. The recent developments in pre-trained language models and their capabilities of zero-short cross-lingual transfer will be explored to develop a multilingual dialogue summarizer. Later it is a hypothesis that this proposed approach can be used as a baseline method for low-resource natural langue tasks.

## 1.8 Research Challenge

Dialogue summarization is a trendy domain on which both large-scale enterprise companies and small companies are focused. Recently Microsoft cognitive azure team has started developing solutions for dialogue summarization.

During the past years, a wide variety of research projects have focused on text summarization, document summarization, news, etc. Yet the limited resources currently available and the complexity reduce the expansion of technology in this domain. Previous studies have explained how real-world dialogues challenge current summarization models (Zhang et al., 2021). Unlike text or document summarization, dialogues carry more complex attributes such as general knowledge, intentions, and informal sentences. Therefore, existing text summarization techniques cannot directly apply to dialogue summarization.

With the limited resources of datasets available in high-resource languages, existing work and future work is more focused on a few languages. This restricts the development of practical solutions that can summarize dialogues other than English. Considering these identified factors, a system to support a multilingual dialogue summarizer is needed.

## 1.9 Chapter Summary

In the introduction chapter author has discussed about the problem with research gap, challenges and the expected contribution to each research and problem domain. Also the research objectives and the learning outcomes are mapped under the requirement of the University of Westminster, BEng(Hons) Software Engineering course final year research project module.

# Chapter 2: Software Requirements Specification

## 2.1. Chapter Overview

This chapter will focus on determining different stakeholders and their interaction to the proposed system, a rich picture diagram and the different requirement gathering methodologies with results. Finally possible use cases with the system, functional and non-functional requirements of the prototype.

## 2.2. Rich Picture Diagram

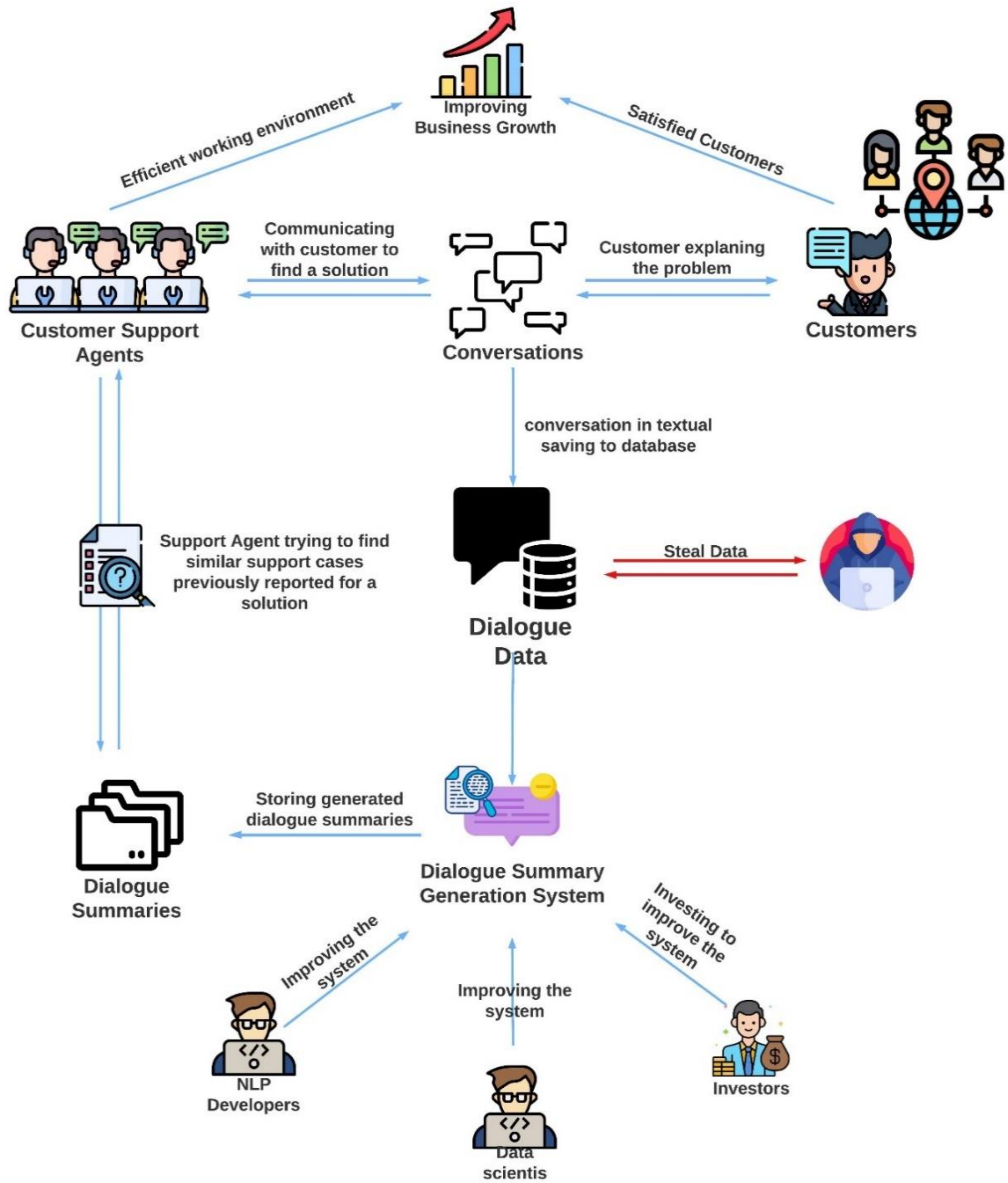



*Figure 2 - Rich Picture Diagram*

## 2.3. Stakeholder Analysis

### 2.3.1 Stakeholder Onion Model

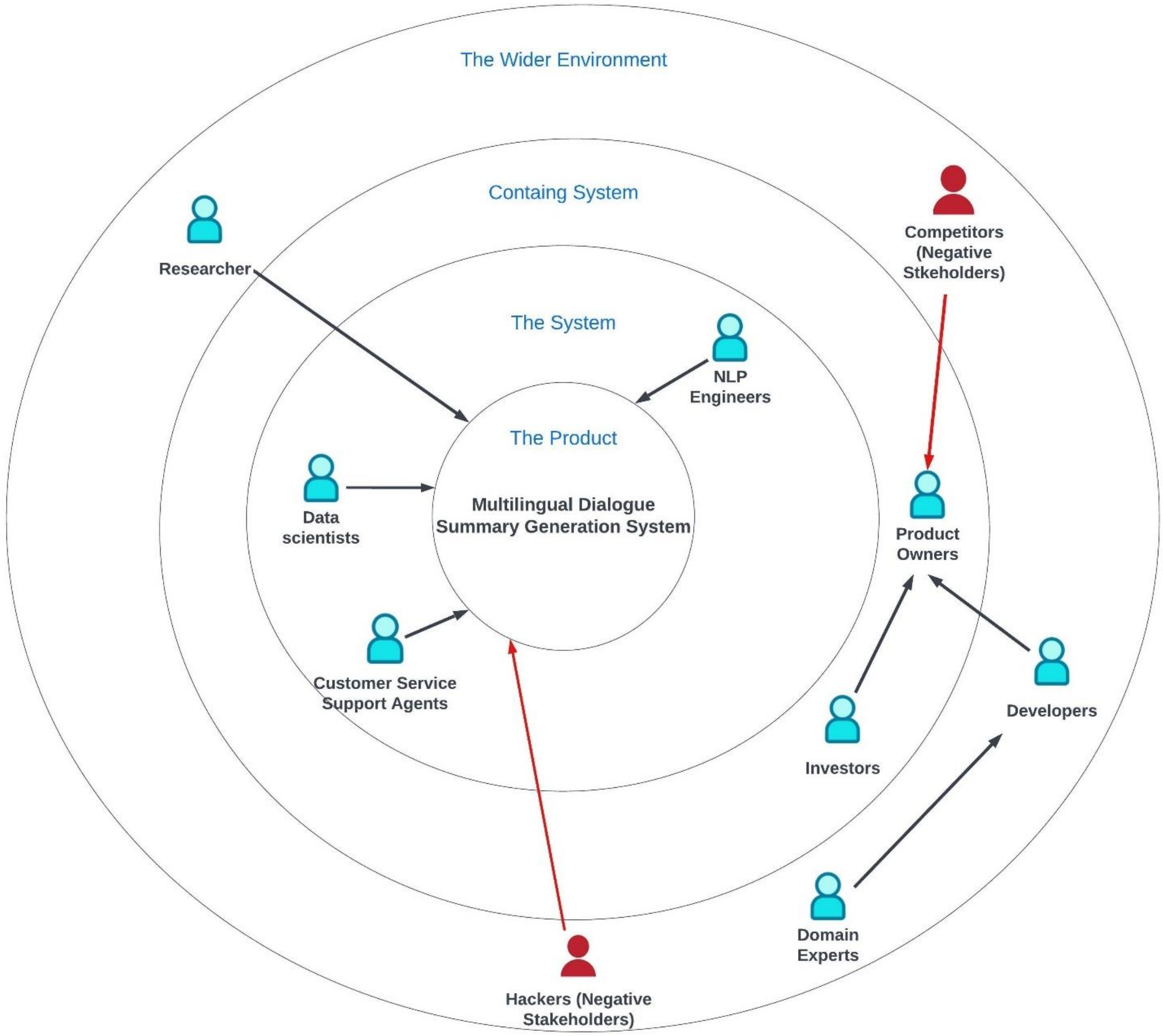



*Figure 3 - Stake Holder Onion Model*

### 2.3.2. Stakeholder Viewpoints

| Stakeholder | Role | Description |
|---|---|---|
| Data scientist, NLP Engineers | Operational Maintenance | Develops the system using tools. |
| Customer Service Support Agent | Functional Beneficiary | Uses the developed system to summarize dialogues. |

| | | |
|---|---|---|
| Product Owners | Functional Beneficiary | Owner of the dialogue summary generation system. |
| Investors | Financial Beneficiary | Invest to make profit and support the future developments and improvements. |
| Competitor | Negative Stakeholder | Creates a system that directly contest with the proposed system's features |
| Hacker | Negative Stakeholder | To disrupt the system and the data it holds. |
| Technical Experts | Expert | Decides whether the product is capable of handling set of requirements from a technical perspective. |
| Researcher | Advisor | Analyze the existing solutions for research purpose. |

*Table 2 - Stake holder Viewpoints*

### 2.3.3. Stake Holder Grouping

| **Stakeholder Groups** | **Stakeholder** | **What is the method that you will be used to gather requirements** |
|---|---|---|
| Group 1 | Customer Service support agents (End-User) | Questionnaires |
| Group 2 | Data scientists and NLP researchers | Interviews |
| Group 3 | Competitors (Existing work) | LR |

*Table 3 - Stakeholder Grouping*

### 2.3.4 Data gathering instruments.

**Questionnaire for Group 1**

| **Question** | **Research/Relevance to the research** | **Research Question** |
|---|---|---|
| Do your organization uses dialogue summarizations? | | RQ1 |
| What is the current technique you are using for dialogue summarization? | Study the currently available solution in the industry. | RQ1 |
| How often are summaries being used for later reference | Impact of dialogue summarization in customer services. | RQ1 |
| Do you summarize all the dialogues? | Identifying the specific attributes that will be considered for dialogue summarization. | RQ1 |
| If you selected "No, only if it is required" for previous questions, select the specific attributes considered to summarize dialogues. | Identifying the specific attributes that will be considered for dialogue summarization. | RQ1 |
| How many languages do you support in customer service? | Overview of language support in customer service. | RQ1, RQ3 |
| Do you find it challenging to use currently available solutions to summarize dialogues that are not in English? | Understanding limitations in current solutions from the end-user's perspective. | RQ1, RQ2 |
| How much do you think it will be helpful to summarize dialogues which are not in English languages? | Identifying the end-user's requirement for the proposed system | RQ2, RQ3 |
| Will you be interested in a solution that can summarize the dialogues between customer support and the customer? | Identifying the end-user's requirement for the proposed system | RQ2, RQ3 |
| If you have any suggestions, please mention them below. | Identifying the end-users' suggestions based on prior experience with existing solutions. | RQ1 |

*Table 4 - Data gathering instruments Group 1*

**Interview Questions for Group 2**

| **Question** | **Research / Relevance to the research** | **Research Question** |
|---|---|---|
| What are the current limitations of using text summarization techniques directly to dialogue summarization | Understanding the current limitations in the dialogue summarization domain. | RQ1, RQ2, RQ3 |
| Recent advancements in cross-lingual transfer techniques in the NLP domain | An exploratory study on cross-lingual transfer techniques. | RQ2 |
| How effective are cross-lingual transfer techniques for low-resource language use cases? | Identifying the practicality of using cross-lingual transfer techniques | RQ2, RQ3 |

*Table 5 - Data gathering instrument Group 2*

## 2.4 Selection of Requirement Elicitation Methodologies

There were multiple requirement elicitation methodologies were followed to gather requirement for this research project. Literature review, interview, survey, and prototyping are selected as for this purpose.

| **Method 1: Literature Reviews** |
|---|
| In the beginning of the project, the author conducted an analysis of the existing research projects related to the chosen domain to identify a solid research gap. The existing system and the technologies were thoroughly studied for this research project. These finding are mentioned in literature. |
| **Method 2: Interviews** |
| Collecting knowledge to the research body and to get qualitative feedback on the proposed solution, to get insights from the domain specific experts several interviews were conducted. By using this method, it supports the researcher to early identify the challenges of the proposed solution when it comes to development of the prototype. |

| **Method 3: Survey** |
|---|
| A questionnaire was used to gather requirements from the end user's perspective. This will help to identify the requirements and the problems with a existing solutions that the end user is facing. These data will be helpful to improve the proposed solution to become more practical. |
| **Method 4: Prototyping** |
| Selected software development life cycle for this research project is agile as it would help to improve the proposed system in a recursive development method. |

*Table 6 - Requirement Elicitation Methodologies*

## 2.5 Discussion of Findings

### 2.5.1 Literature Reviews

| Finding | Citation |
|---|---|
| The proposed method shows that latent variables cope with the variance of semantically similar sentences across different languages. Cross-lingual transfer between English to Spanish and English-to-Thai have demonstrated state-of-the-art results. | (Xiang et al., 2021) |
| The proposed framework can generate an abstractive summarization with limited parallel resources by sharing a unified decoder that generates both monolingual and cross-lingual summaries | (Bai, Gao and Huang, 2021) |
| This proposed approach has shown the state-of-the-art models are very sensitive to language shift through automatic translation and combining training data for the two languages (English – Italian) is beneficial. | (Labruna and Magnini, 2021) |
| This proposed method uses a novel Leader-writer network with auxiliary key point sequences, ensuring the generated summary is logical and integral. | (Liu et al., 2019) |

*Table 7 - Discussion of Findigs LR*

### 2.5.2 Interviews

To get opinion on the proposed solutions and the research area both research experts from dialogue summarization and domain experts from the customer service area were chosen. The interviews were conducted as open-ended question therefore the output of these interviews was documented based on thematic analysis.

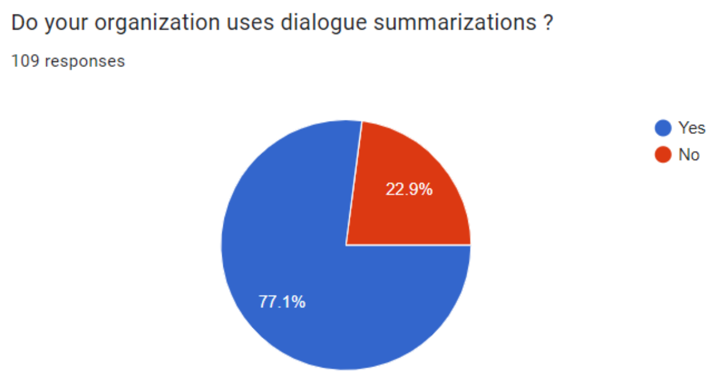
| **Codes** | **Theme** | **Conclusion** |
|---|---|---|
| • Slow-moving<br>• Utilized existing resources.<br>• Missing context | Existing limitations in dialogue summarization domain. | All the participants were mentioned about the current text summarization techniques and why it cannot directly apply to the dialogue summarizations. Because the nature of the dialogues constructive and logical text data, keeping the context while summarizing is challenging compared to text summarization. According to the experts they suggest to fine tune text summarization techniques and use of pre-trained language models to build the base stage for dialogue summarizations rather than developing from the scratch. |
| • Not well explored.<br>• Transfer learning | Integrating cross-lingual transfer techniques for low resources language. | Most of the cross-lingual transfer techniques are not yet explored in the dialogue summarization domain. Due to the minimum number of datasets that are currently available, use of cross-lingual transfer techniques will be very useful. Experts suggest to use of a cross-lingual transfer modal which can be used as a intermediate process for machine translation when developing the solution. |

| | | |
|---|---|---|
| • Research gap<br>• Contribution | Research gap and Scope | The opinion from the research domain experts were that the research gap and the proposed solution is an innovative use of existing methods to overcome the problem. |
| • Globalized customer base.<br>• Human resources.<br>• Practicality | Understanding the use of dialogue summarization in a practical use case. | As from the experts from the problem domain, they suggested that the use of supporting to more languages is very useful as one of the problems they are facing is the human resources cost that is required when it comes to manually summarizing the dialogues. Throughout the interviews the customer supports that have globally expanded have this as a major issue when comparing with services that are only available in English. |
| • Save work.<br>• Later refer.<br>• Generate Insights | Features of the proposed system. | Proposed system should be capable of saving the generated summaries and allow the users or the system administrators to refer later. As from the experts they mentioned that summaries can be use for various purposes later such as categorize and analyze to get new insights related to customers issues. |

*Table 8 - Discussion of Findigs Interview*

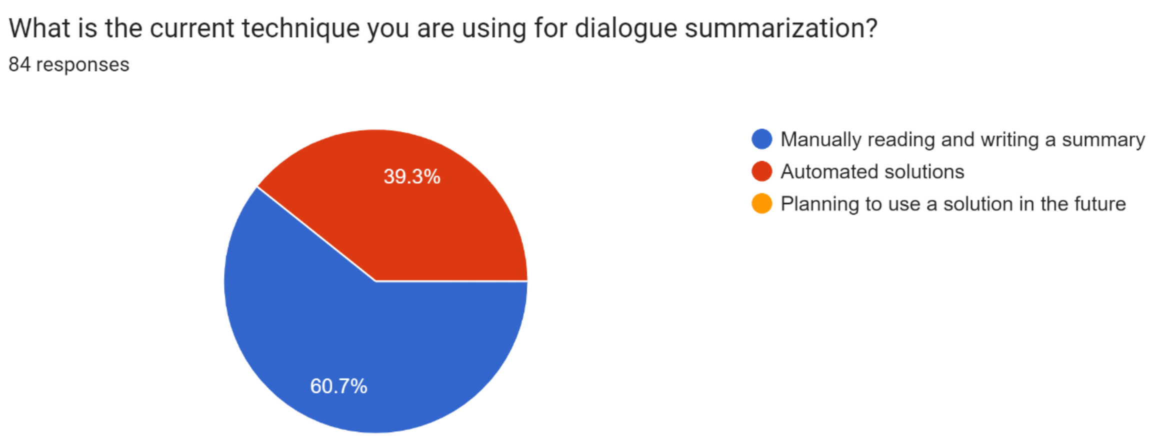
### 2.5.3 Survey

| | |
|---|---|
| Question | Do your organization uses dialogue summarizations? |
| Aim of Question | To understand the awareness of these tools among the end users. |



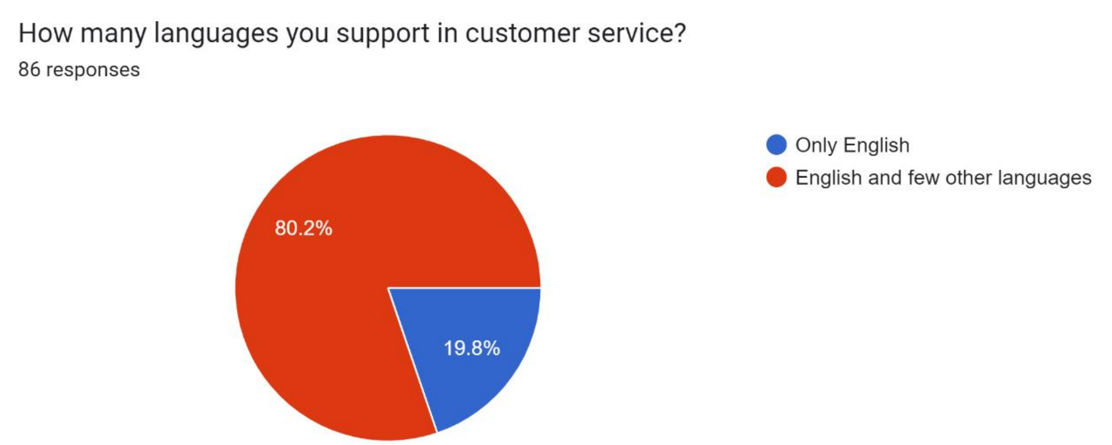

Majority of the participants are aware and currently using dialogue summarization within their organizations. Lesser number of responses shows that not all the customer services are not using dialogue summarizations.

| | |
|---|---|
| Question | What is the current technique you are using for dialogue summarization? |
| Aim of Question | Study the currently available solution in the industry. |



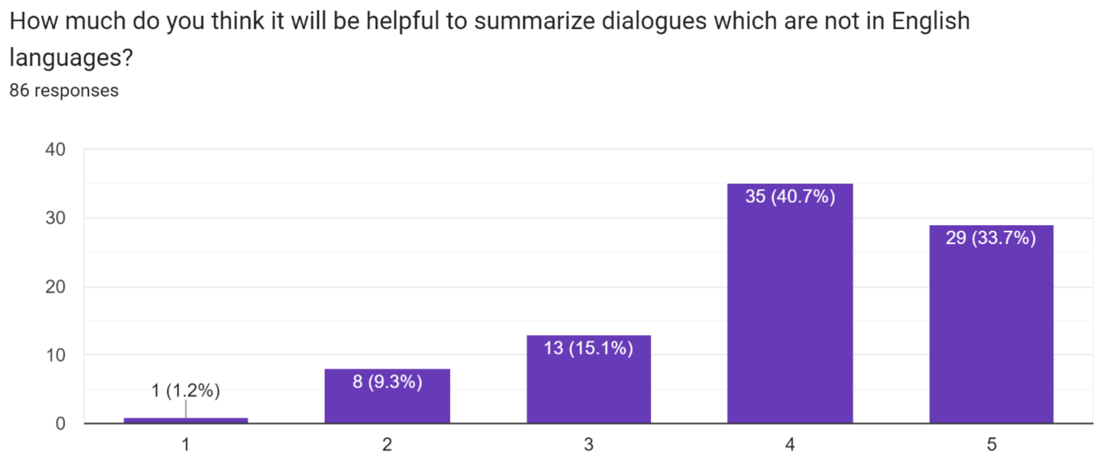
| |
|---|
| From the gathered responses it clearly shows that majority of the users are summarizing dialogues by manually. 39.3% of users are already using automated solutions for dialogues summarization. Observation from the this can be considered that the awareness of the dialogue summarization solutions are not yet expanded in the industry. This can be a major reason that the tools are not yet well developed or in the early stages. |

| | |
|---|---|
| Question | How many languages you support in customer service? |
| Aim of Question | Overview of language support in customer service. |




Majority of the customer services supports more than English language. This can be potential requirement as dialogue summarization that it should not be limited for English.

| | |
|---|---|
| Question | How much do you think it will be helpful to summarize dialogues which are not in English languages? |
| Aim of Question | Identifying the end-user's requirement for the proposed system |




From the responses that more than 60% of users are interested in the proposed solution which can be helping to solve the existing problems with the current solutions or to replace the current solutions.

*Table 9 -Discussion of Findigs Survey*

### 2.5.4 Prototyping

| Criteria | Findings |
|---|---|
| Developing core component | Use of transformers and their capabilities needs to be identified clearly. Different transformers have different capabilities |
| Applying cross-lingual transfer techniques | Final output of the summary must be clear. Context cannot be changed during the transfer process as it may lead to different meaning. |

*Table 10 - Prototyping*

### 2.5.4 Summary of Findings

| Finding | Literature Review | Interviews | Questionnaire | Prototyping |
|---|---|---|---|---|
| The proposed system will help to benefit users who are using existing | | ✓ | ✓ | ✓ |

| | | | | |
|---|---|---|---|---|
| solutions and users with no prior use of dialogue summarization solutions. | | | | |
| The limitation in dialogue summarization systems can be pushed by using pre-trained language models and cross lingual transfer techniques. | ✓ | ✓ | ✓ | |
| Identified research gap would contribute to dialogue summarization research domain. | ✓ | ✓ | | |
| Within the system summaries should be stored and allow user or administrators for later reference. | | ✓ | | |

*Table 11 - Summary of Findings*

## 2.6 Context Diagram

The scope of the proposed system and the interaction with internal and external components should be identified before the development. The following diagram represent the context of the proposed system.

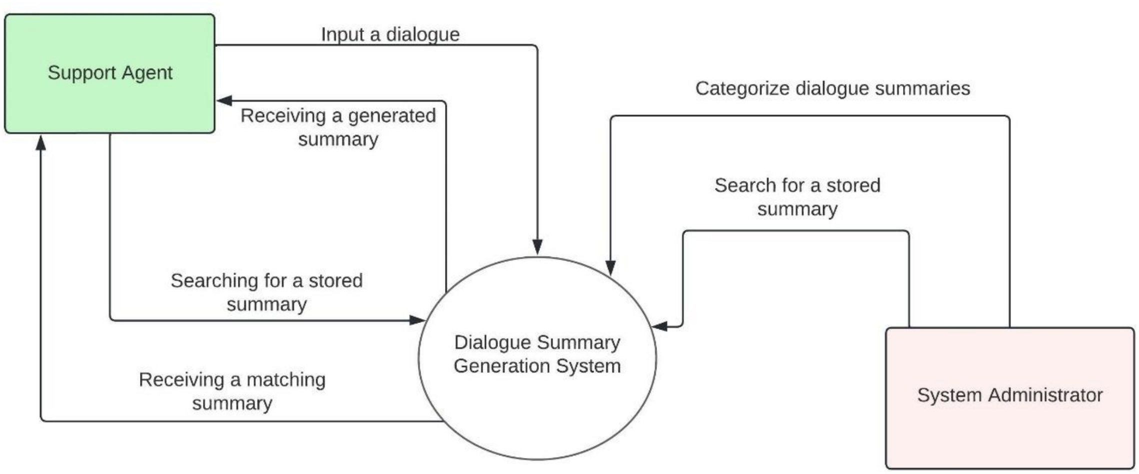



*Figure 4 - Context Diagram*

## 2.7 Use case Diagram

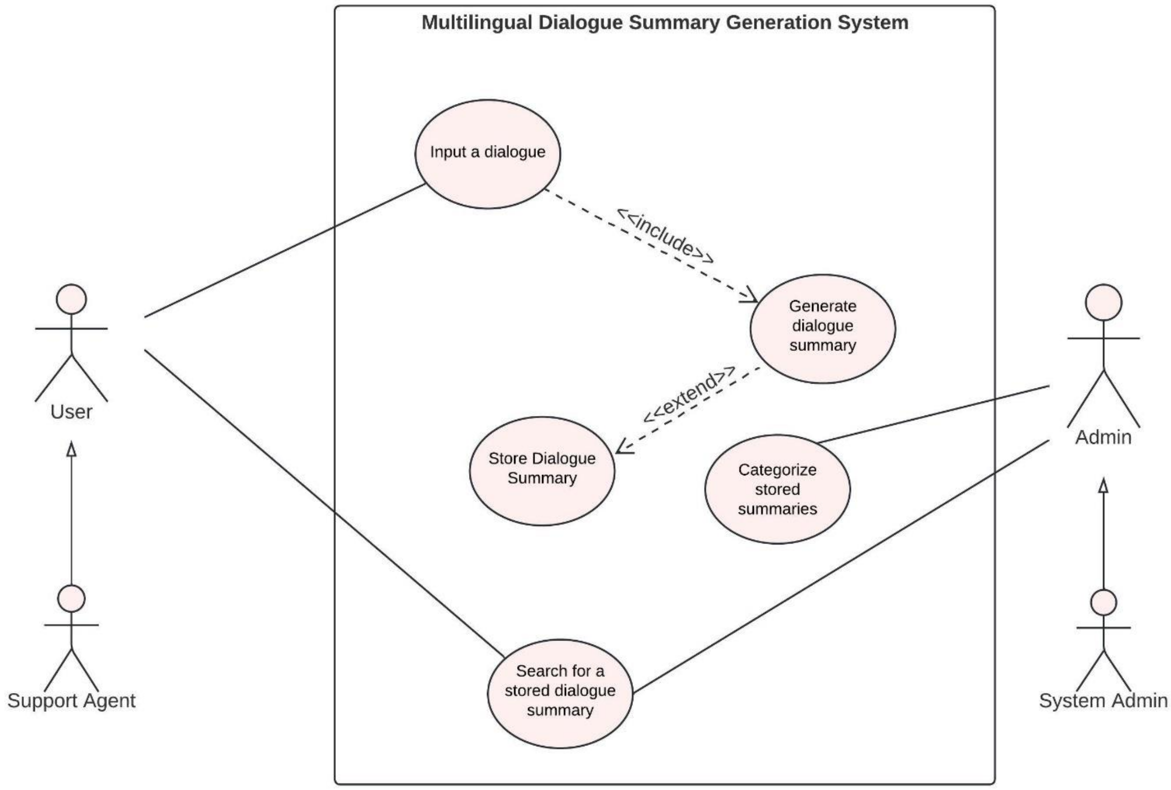



*Figure 5 - Use Case Diagram*

## 2.8 Use case description

| Use Case: UC1 | Generate Dialogue Summary |
|---|---|
| Description | Generate a summary for the dialogue input |
| Primary Actor | Support Agent (User) |
| Pre-condition | Textual dialogue data should be in a valid input format. |
| Post condition | Generated summary should be displayed to the user. |
| Trigger | A support agent wishes to generate a summary for the dialogue which he/she had with the customer. |

| | |
|---|---|
| Success scenario | • A summary for the input data will be generated.<br>• Generated summary will be stored in the system. |

*Table 12 - Use Case: UC1*

| Use Case: UC2 | Search for a stored dialogue summary |
|---|---|
| Description | Find a previously generated summary with matching search results. |
| Primary Actor | Support Agent (User), System Admin |
| Pre-condition | Previously generated summaries should be saved on the system. |
| Post condition | Matching summaries should be displayed. |
| Trigger | A support agent / system admin wishes to search for previously generated summaries. |
| Success scenario | • Matching summaries will be displayed.<br>• If there are no matching summaries will display a message as 0 results. |

*Table 13 - Use Case: UC2*

| Use Case: UC3 | Categorize stored summaries |
|---|---|
| Description | Categorize summaries based on their similarities. |
| Primary Actor | System Admin |
| Pre-condition | Previously generated summaries should be saved on the system. |
| Post condition | Summaries should be categorized. |
| Trigger | A system admin wishes to categorize summaries based on their similarities to generate insights. |
| Success scenario | • Stored summaries will be categorize based on their similarities. |

*Table 14 - Use Case: UC3*

## 2.9 Requirements

### 2.9.1 Functional Requirements

In order to prioritize the levels of the system requirements the MoSCoW principle was used.

| **Priority Level** | **Description** |
|---|---|
| Must have (M) | This priority level indicates the core functional requirements and must be implemented. |
| Should have (S) | This priority level indicates the requirements that are not necessary but do add a great value. |
| Could have (C) | This priority level indicates the optional requirements |
| Will not have (W) | This priority level indicates the requirements that are out of the project scope. |

*Table 15 - MoSCoW Priority*

| **FR ID** | **Requirement** | **Priority Level** | **Use Case** |
|---|---|---|---|
| 1 | User must be able to generate a summary for an inputted dialogue data at least for 1 language option. | M | UC1 |
| 2 | Summary of the dialogue should be represented to the user. | M | UC1 |
| 3 | User should be able to generate summary for a dialogue data with multiple language options. | S | UC1 |
| 4 | The system should store the generated dialogue summaries and allow users to view the summaries | S | UC2 |
| 5 | The system should allow the users to search stored summaries with matching keywords. | S | UC2 |
| 6 | The system should allow the admins to categorize the stored summaries based on their similarities | C | UC3 |

*Table 16 - Functional Requirement*

### 2.9.2 Non-functional requirements

| NFR ID | Requirement | Description | Priority Level |
|---|---|---|---|
| 1 | Quality of the Output | The quality of the generated summary should be clear and meaningful as much as possible. | M |
| 2 | Performance | Time to generate a summary should be acceptable. | S |
| 3 | Security | The system should prevent any data breaches from attackers to keep the information safe. | S |
| 4 | Usability | The system should be easy to use followed by good user interface and user experience principles. | M |

*Table 17 - Non-functional Requirement*

## 2.10 Chapter Summary

In this chapter, to represent the overview of the system and how it connects with the different parties have been displayed using the rich picture diagram. Sounder’s onion model was used to display the stakeholders and how they connect with each other. In order to get opinions for the proposed solution different requirement elicitation methodologies were followed using the stakeholders. In the latter part the use cases of the system, functional and nonfunctional requirements were documented based on the input from the data gathering results.

# Chapter 3: Design

## 3.1 Chapter Overview

This chapter will include the all the design decision that were made to develop the proposed system. Authors main design goals, overview of the architecture and both high-level and low-level design and wireframes will be discussed.

## 3.2 Design Goals

| Design Goals | Description |
|---|---|
| Correctness | The output summary should be accurate and should not drop the context of the dialogues. The proposed system should be able to summarize the dialogues with similar to human annotated summaries. |
| Performance | The system should be capable of handling lengthy dialogues and optimized to generate summaries within a short period of time. |
| Usability | One of goal of the proposed system is to minimize the human resources that will take to summarize the dialogues manually, so it should be user friendly and be able to work with minimum effort. |
| Testability | The system should be divided into components within the development process to test and identify errors in the early stages. |
| Scalability | In the production environment there will be more users using the system and the workload should be handled to continue the process. |

*Table 18 - Design Goals*

## 3.3. High-Level Design

### 3.3.1 Tiered Architecture

The overview of the system architecture is given bellow, in the tiered architecture diagram. The presentation, logic and the data layers are organized in the three-tier architecture.

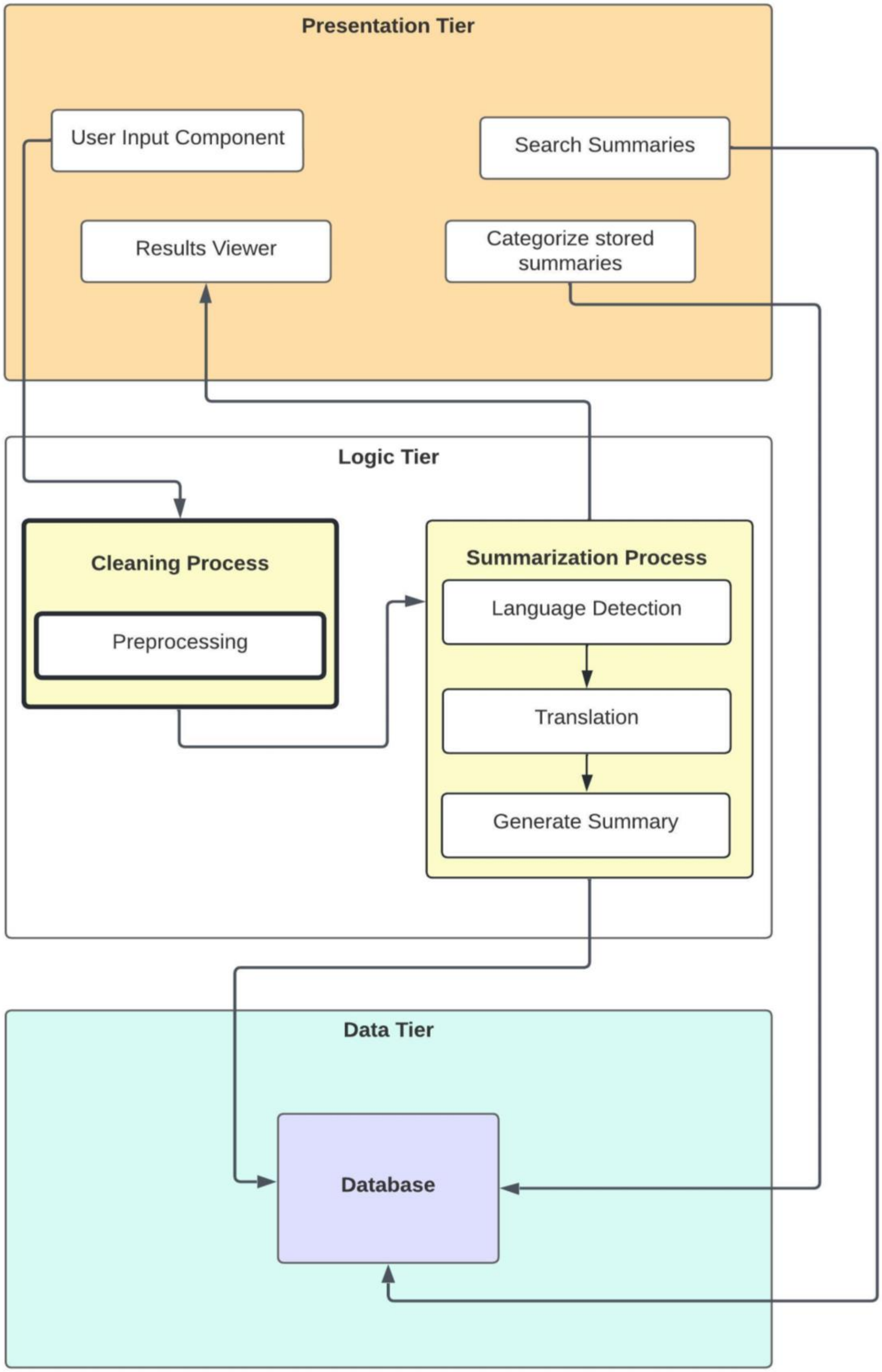



*Figure 6 - Tiered Architecture Diagram*

Above composed diagram is represented in a modular form which will be easy to understand, apart from the above diagram a backend services will be used to communicate among each layer.

### 3.3.2 Discussion of tiers

**Data Tier**

Data tier will consist of the database which will be using to store the generated dialogue summaries. Also, the user logins will use the database for authentication purposes to ensure the system security.

**Logic Tier**

Logic tier will be responsible for processing the user input before putting through the summarization process. User input will be pre-processed by removing stop words etc. Then the processed input will be pass through the language detection modular. Machine translation will modular will go through process where it translates the input while minimizing the loss of context from the original user input. Then finally the summary will be generated.

**Presentation Tier**

User input component is where the end-user will input the dialogue in textual format. Results viewer will display the generated summary to the user. Search dialogue summaries will enable the user to search through previously generated summaries with matching keywords from the database.

## 3.4. System Design

### 3.4.1. Choice of the Design Paradigm

After the requirements gathering through multiple requirement elicitation methodologies Author has selected the **SSADM** (Structured Systems Analysis and Design Method) for the development of the proposed prototype because the requirement for the proposed system is well defined.

## 3.5. Design Diagrams

### 3.5.1 Data Flow Diagram

#### Level 01

Level 1 Data flow diagram represent the flow of data through different process within the system.

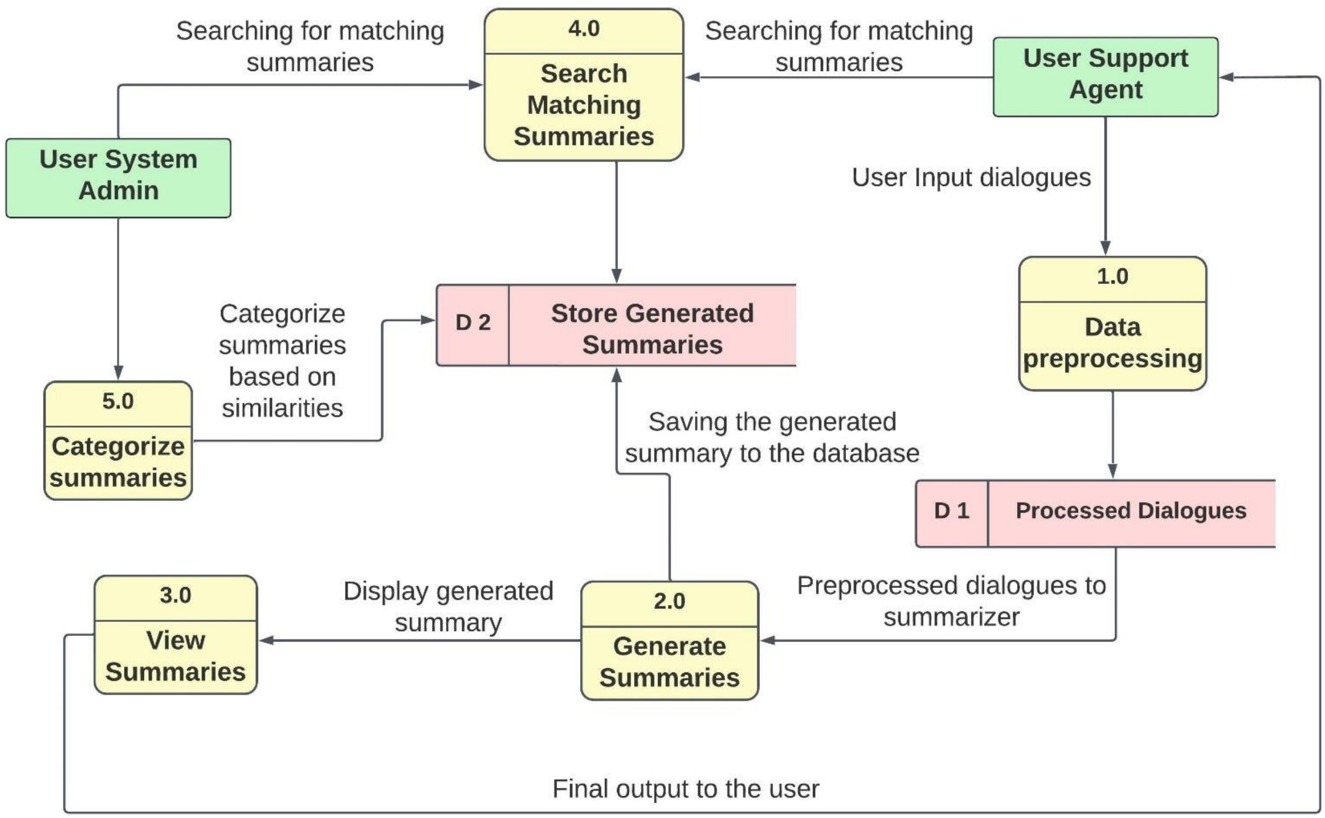



*Figure 7 - Data Flow Diagram Level 1*

## Level 02

Level 2 Data flow diagram represent more expanded breakdown of the level 1 data flow diagram.

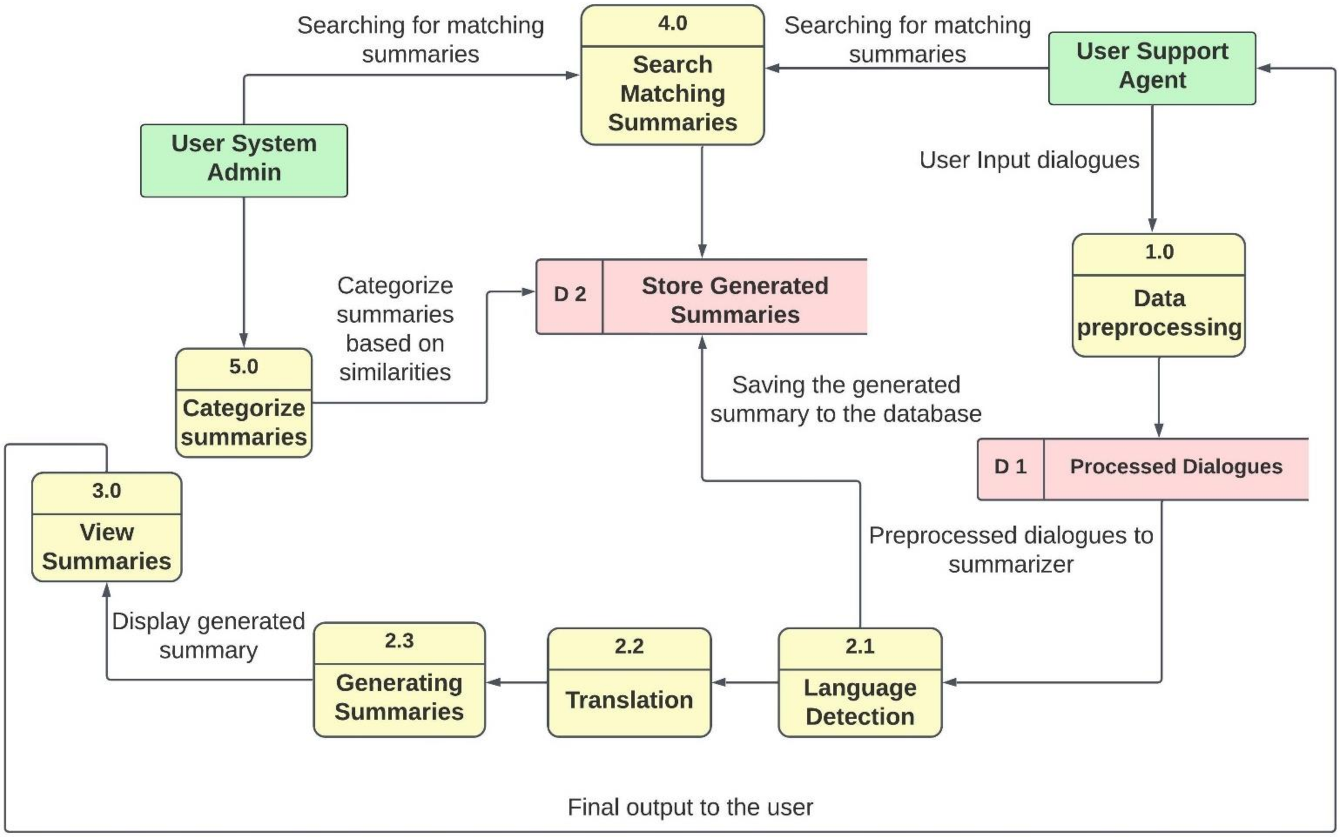



*Figure 8 - Data Flow Diagram Level 2*

### 3.5.2 System Process Flow Chart

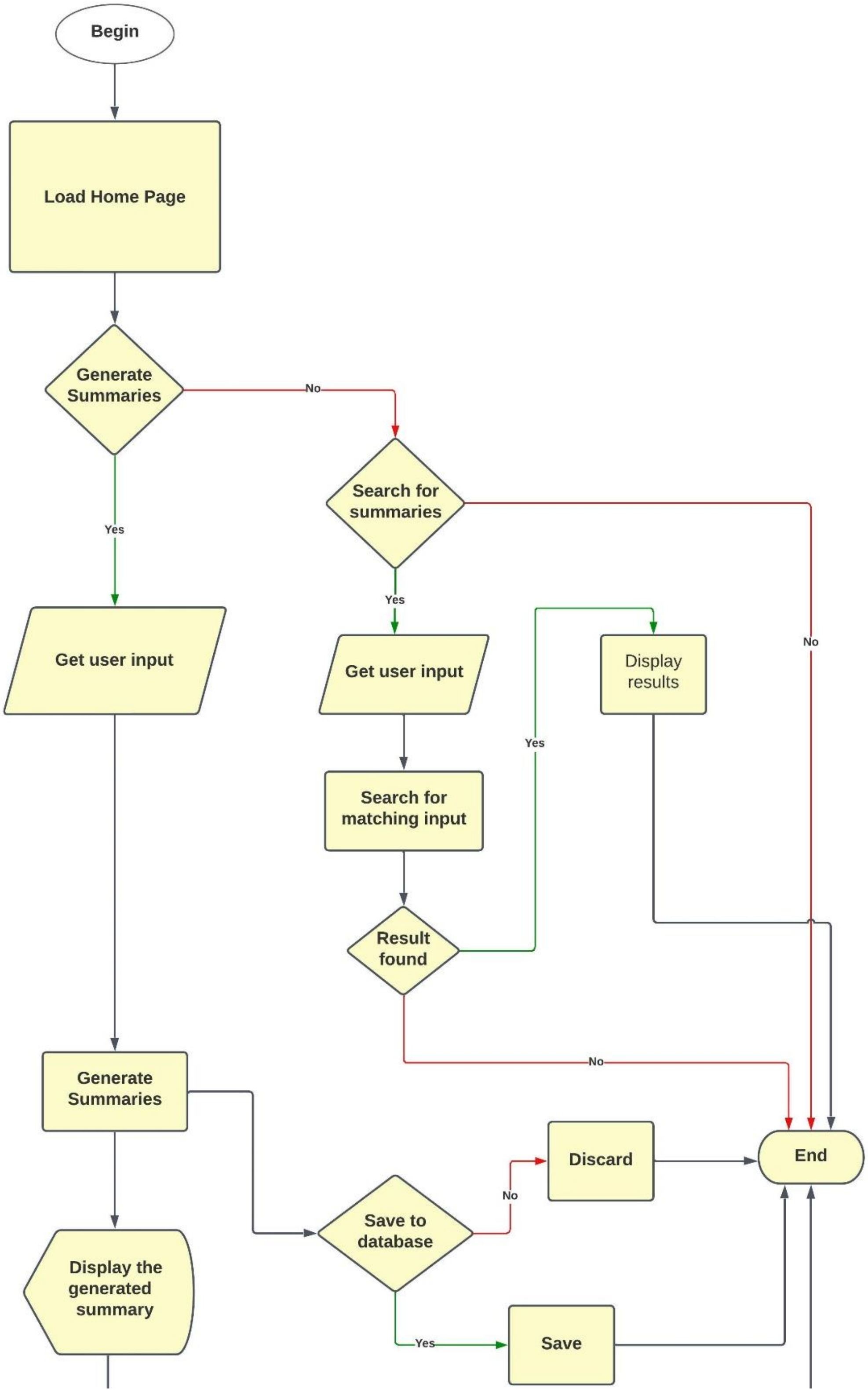



*Figure 9 - System Process Flow Chart*

### 3.5.3 User Interface Design

### 1. Low level fidelity wireframe diagram


*Figure 10 - Low level fidelity wireframe*

**2. High level fidelity prototype**

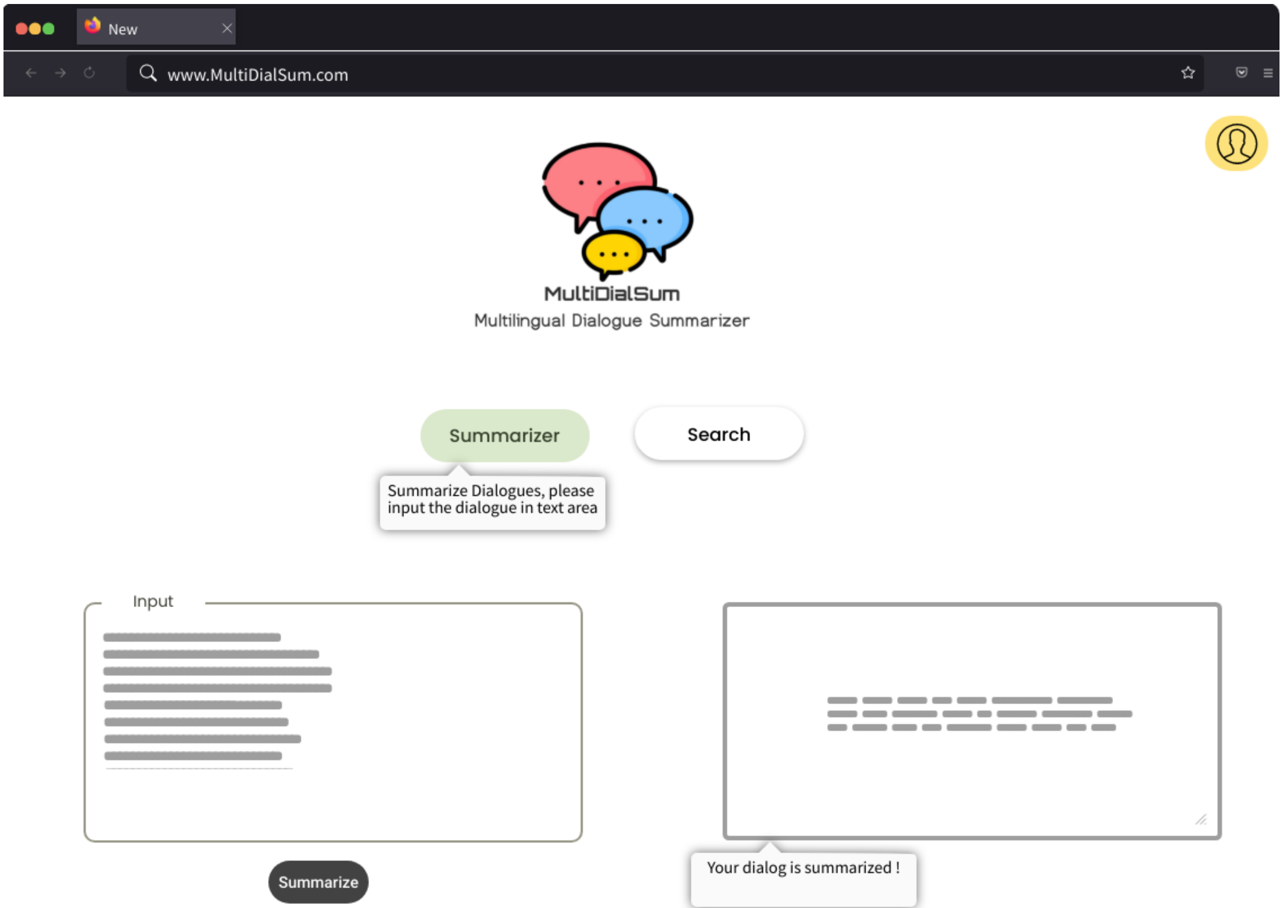



*Figure 11 - High level fidelity wireframe*

## 3.6. Chapter Summary

Design of the proposed system and the architecture of the system were discussed, also the overview of the expected UI for the system is explained using wireframe diagrams in this chapter.

# Chapter 4: Initial Implementation

## 4.1 Chapter Overview

This chapter will consist of the technology stack, data set selection and other information related to the development of the prototype. Reasons for selecting those specific attributes will be discussed.

## 4.2 Technology Selection

### 4.2.1 Technology Stack

The chosen technologies for the proposed system in each layer is represented in the bellow diagram.

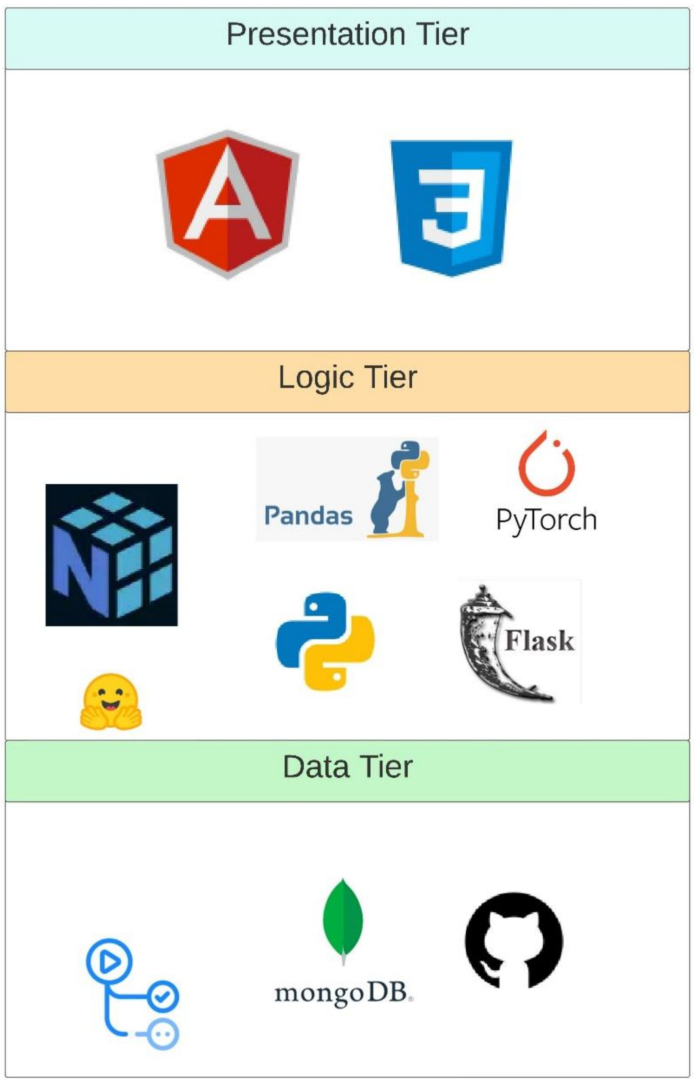



*Figure 12 - Technology Selection*

### 4.2.2 Data-set Selection

This research project requires a dataset with dialogues in customer service domain. As for the data set selection TWEETSUM dataset was chosen for the project. Tweetsum dataset is focused on twitter customer care conversation. It was made with 1,100 dialogs from tweets that is available on the Kaggle customer support on twitter. Each dialogue is summarized with 3 extractive and 3 abstractive summaries by human annotators. Tweetsum dataset contains total summaries of 6500. (Feigenblat et al., 2021). Dataset is publicly available on the official GitHub repo under CDLA-sharing license.

### 4.2.3 Development Framework

| Framework | Justification for selection |
|---|---|
| Flask | Since the proposed system is a web application, flask is chosen. Flask is a light weight and extensible web framework that will allow to develop web applications. Also, flask has integrated unit testing feature which will be useful for API test validation. |
| Material Design | Material design is used by many industrial applications. Material design is useful for rapid prototyping as it will allow the developer to focus on the logical components while not spending too much time on the css and other ui parts. |
| PyTorch | Main reason to choose pytorch over its competitive machine learning framework tensorflow is that pytorch well-designed comparing to tensorflows frequent api changes. |

*Table 19 - Development Framework*

### 4.2.4 Programming Languages

Python will be used as the main development language for the core component to preprocess the dataset and training modals. Python is one of the most chosen programming languages for data science projects because of its package’s availability and support.

For the API development the flask framework is chosen, and the programming language will be python. Flask is a python based microframework which will allow to develop light weight web APIs.

For the frontend development, typescript was decided to be used. Even typescript will require more time to compile the code when comparing with the plain JavaScript, typescript allows to check the correctness at the compile time which will be useful while developing the prototype.

### 4.2.5 Libraries

| **Library** | **Justification for the selection** |
|---|---|
| Pandas | Pandas dataframe library is capable of performing various operation on datasets such as cleaning, sorting, filtering etc. |
| Numpy | Numpy python library allow to perform mathematical operation on multidimensional arrays. |
| transformers | Transformers libraries provide API access to tools with large amount of pre-trained language transformer modals. |
| ROUGE | This rouge python package will allow to access the rouge metrices for evaluating the summarization and machine translations. |
| TensorBoard | TensorBoard is a package that will allow to visualize the machine learning related work such as loss and accuracu etc. TensorBoard can be easily integrate with pytorch. |

*Table 20 - Libraries*

### 4.2.6 IDE

| **IDE** | **Justification for the selection** |
|---|---|
| Google Collab | More convenient cloud-based development environment and being able to work on multiple devices. |
| PyCharm | Fully fledged IDE which support lot of useful tools for local development. |
| VS code | Convenient IDE for development with lot of tools and extensions. |

*Table 21 - IDE Selection*

### 4.2.7 Summary of Technology Selection

| **Component** | **Tools** |
|---|---|
| Programming language | Python, typescript |
| Development framework | Flask |
| UI framework | Material design of Angular |
| Libraries | Pandas, Numpy, transformers, ROUGE, TensorBoard |
| IDE | Google Collab (core functionality), PyCharm (API development), VS code (frontend) |
| Version control | Git, GitHub, Hugging Face |
| Web app hosting | Netlify (frontend-host), AWS (backend-host) |
| CI/CD | Github Actions |

*Table 22 - Summary of the Technology Selection*

## 4.3 Implementation of the Core Functionality

For the implementation, the format of the tweetsum data set needs to be changed. Summaries in the tweetsum dataset is annotated by humans. But the tweetsum dataset doesn't have original conversations instead it has tweet Ids. So for this research project will be using the twitter customer support data and tweetsum data set to map the conversations and their respective summaries into a single data set for the modal training purposes. Following code is to extract those conversations and summaries from both datasets to finally produce a single complete dataset.

```
!pip install pandas
!pip install json
!pip install re
```

```
[ ] import pandas as pd
    import json
    import re
    import numpy as np
    import re

    # Reading main customer support dataset from kaggle
    df_twcs = pd.read_csv('drive/MyDrive/ColabNotebooks/Dataset/twcs.csv')
```

```
[ ] # opening jsonl files to map with main customer support data
    with open('drive/MyDrive/ColabNotebooks/Dataset/tweet_sum_data_files/final_test_tweetsum.jsonl') as f:
        lines = f.read().splitlines()
        df_tweetsum_json = pd.DataFrame(lines)
        df_tweetsum_json.columns = ['json_element']
```

```
[ ] # preparing the TWEETSUM dataset files to a csv files with mapped coloumns
    df_tweetsum_json['json_element'].apply(json.loads)
    df_tweetsum_csv = pd.json_normalize(df_tweetsum_json['json_element'].apply(json.loads))
    df_tweetsum_csv.to_csv('tweetsum_test.csv') # tweetsum json to csv

    list_tweet_id = [] # python list to append rows with tweet id
    tweet_id_offset = 1 # tweet_id_offset coloumn in df_tweetsum_csv (tweetsum_test.csv)
    for index, col in df_tweetsum_csv.iterrows():
        arr = col[tweet_id_offset]
        df_row = pd.DataFrame(arr)

        # adding to a list
        list_tweet_id.append(df_row['tweet_id'])
    # writing to a csv file with tweet id sequnce
    df_final_output = pd.DataFrame(list_tweet_id)
    print(df_final_output)
    df_final_output.to_csv('tweet_id_seq.csv')
```

*Figure 13 - Implementation of the core function 1*

```
# loading dataset with tweet id sequnce
df_preprocessed = pd.read_csv('tweet_id_seq.csv')

tweet_id_seq_arr  = [] # array to append tweet id
arr = df_preprocessed.to_numpy() # converting the dataframe to a numpy array with float numbers and NaN values

# function to remove float numbers
def formatNumber(num):
  if num % 1 == 0:
    return int(num)
  else:
    return num

# removing tweet_id text, and NaN values from the numpyarray
for i in range (df_preprocessed.shape[0]):
  index = np.argwhere(arr[i]=='tweet_id')
  y = np.delete(arr[i], index)
  z = y.tolist()
  cleanedList = [x for x in z if str(x) != 'nan']
  for i in range(len(cleanedList)):
    cleanedList[i] = formatNumber(cleanedList[i])
  tweet_id_seq_arr.append(cleanedList)


print(tweet_id_seq_arr)
```

*Figure 14 - Implementation of the core function 2*

```
# Creating dataset with matching tweet Ids
#   **dialogue | summary** format

#print(df_twcs)

#print(tweet_id_seq_arr[0])
#df_twcs.loc[df_twcs['tweet_id'] == 1]
text_arr = []
for tweet_id in tweet_id_seq_arr[0]:
  text = df_twcs['text'].loc[df_twcs['tweet_id'] == tweet_id].values
  text_arr.append(text)

# 0, 2, 4 ...
def customer_dialog(text):
  first = text.partition(' ')[0]
  customer_txt = text.replace(first, 'customer:')
  return customer_txt
# 1, 3, 5 ...
def support_dialog(text):
  first = text.partition(' ')[0]
  customer_txt = text.replace(first, 'support:')
  return customer_txt
```

*Figure 15 - Implementation of the core function 3*

Next process will be to train the transformer backbone modal which will be the ‘bart’ using the processed dataset. For the modal training will be using the hugging face trainer API. With this API modal training will be done using a hugging face server machine.

```
import transformers
from transformers import AutoTokenizer, AutoModelForSeq2SeqLM, Seq2SeqTrainingArguments, Seq2SeqTrainer
from datasets import load_dataset, load_from_disk
import numpy as np
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

```
max_input = 512
max_target = 128
batch_size = 3
model_checkpoints = "facebook/bart-large-xsum"
data = load_dataset("tweetsum_processed.csv")
```

```
tokenizer = AutoTokenizer.from_pretrained(model_checkpoints)
```

```
def preprocess_data(data_to_process):
  #get all the dialogues
  inputs = [dialogue for dialogue in data_to_process['dialogue']]
  #tokenize the dialogues
  model_inputs = tokenizer(inputs,  max_length=max_input, padding='max_length', truncation=True)
  #tokenize the summaries
  with tokenizer.as_target_tokenizer():
    targets = tokenizer(data_to_process['summary'], max_length=max_target, padding='max_length', truncation=True)

  #set labels
  model_inputs['labels'] = targets['input_ids']
  #return the tokenized data
  #input_ids, attention_mask and labels
  return model_inputs
```

```
[ ] tokenize_data = data.map(preprocess_data, batched = True)
```

```
args = Seq2SeqTrainingArguments(
    'finetuned-multidial', #save directory
    evaluation_strategy='epoch',
    learning_rate=2e-5,
    per_device_train_batch_size=2,
    per_device_eval_batch_size= 2,
    gradient_accumulation_steps=2,
    weight_decay=0.01,
    save_total_limit=2,
    num_train_epochs=3,
    predict_with_generate=True,
    eval_accumulation_steps=3,
    fp16=True #available only with CUDA
    )
```

```
trainer = Seq2SeqTrainer(
    model,
    args,
    train_dataset=tokenize_data['train'],
    eval_dataset=tokenize_data['validation'],
    tokenizer=tokenizer
)
```

```
[ ] trainer.train()
```

*Figure 16 - Implementation of the core function 4*

After the modal training completion, the modal will be saved to personal hugging face account, which will be useful for directly calling the modal.

```
!pip install transformers
from transformers import pipeline
summarizer = pipeline("summarization", model="mr-desilva/finetuned-multidial") # calling the saved modal

# Sample Test
conversation_tweetSum = '''customer:My watchlist is not updating with new episodes (past couple days).  Any idea why?
support:Apologies for the trouble, Norlene! We're looking into this. In the meantime, try navigating to the season / episode manually.
customer:Tried logging out/back in, that didn't help
support:Sorry! 😔 We assure you that our team is working hard to investigate, and we hope to have a fix ready soon!
customer:Thank you! Some shows updated overnight, but others did not...
support:We definitely understand, Norlene. For now, we recommend checking the show page for these shows as the new eps will be there
customer:As of this morning, the problem seems to be resolved. Watchlist updated overnight with all new episodes. Thank you for your attention t
support:Awesome! That's what we love to hear. If you happen to need anything else, we'll be here to support! 💚'''

summary = summarizer(conversation_tweetSum)

print(summary)
```

*Figure 17 - Implementation of the core function 5*

## 4.4 User Interface

Current development stage doesn't have a proper frontend development. All the designs related wireframe and mockups are included in the design chapter.

## 4.5 Chapter Summary

This chapter is consisting of the technologies and tools that were selected to the development of the proposed system with the reasonings. Also, the discussion on the development of the core component.

# Chapter 05: Conclusion

## 5.1 Chapter Overview

This chapter is consisting of the conclusion of the product specification and prototype design. Deviations with both scopes related, and schedule related, initial test results and required improvements for the final product.

## 5.2 Deviations

### 5.2.1 Scope related deviations.

As for the status of the project, there are no scope deviations from the project proposal.

### 5.2.2 Schedule related deviations

There are no schedule-related deviations in progress. Fine-tuning of the prototype (Minimum viable product) to align with the Gantt chart is expected to be delivered as planned in March 2023.

## 5.3 Initial Test Results

After running 3 epochs the modal scored training loss of 0.144 and validation loss of 0.35. This will be fine-tuned and evaluate with evaluate metrices in the future.

```
***** Running training *****
  Num examples = 14732
  Num Epochs = 3
  Instantaneous batch size per device = 2
  Total train batch size (w. parallel, distributed & accumulation) = 4
  Gradient Accumulation steps = 2
  Total optimization steps = 11049
  Number of trainable parameters = 406290432
```

[11049/11049 1:55:54, Epoch 3/3]

| Epoch | Training Loss | Validation Loss |
|---|---|---|
| 1 | 0.307300 | 0.329423 |
| 2 | 0.199000 | 0.337761 |
| 3 | 0.144500 | 0.352294 |

*Figure 18 - Initial Test Results*

**Output of the dialogue summary generated by the modal.**

```
customer: My watchlist is not updating with new episodes (past couple days).  Any idea why?
support: Apologies for the trouble, Norlene! We're looking into this. In the meantime, try navigating to the season /
episode manually.
customer: Tried logging out/back in, that didn’t help
support: Sorry! 😊 We assure you that our team is working hard to investigate, and we hope to have a fix ready soon!
customer: Thank you! Some shows updated overnight, but others did not...
support: We definitely understand, Norlene. For now, we recommend checking the show page for these shows as the new
eps will be there
customer: As of this morning, the problem seems to be resolved. Watchlist updated overnight with all new episodes.
Thank you for your attention to this matter! I love Hulu ♡
support: Awesome! That's what we love to hear. If you happen to need anything else, we'll be here to support! ♡

Generated Summary -
        "Customer's Hulu watchlist is not updating with new episodes. Some shows updated overnight, but others did
not. The problem has been resolved as of this morning. The watchlist updated overnight with all new episodes, so the
new episodes will be there. "
```

*Figure 19 - Final Output*

## 5.4 Required Improvements

Language translation capability to the summarizer needs to be implemented and overall performance of the summarization and the translation needs to be evaluated with evaluation metrices.

## 5.5 Demo of the Prototype

https://youtu.be/CHCC3ujmIwM

## 5.6 Chapter Summary

This chapter concludes the conclusion of the product specification and prototyping design with a deviation and the required improvements and the link to the demo of the prototype.

# Reference

Chen, Y. et al. (2021). DIALOGSUM: A Real-Life Scenario Dialogue Summarization Dataset. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 5062–5074. Available from https://doi.org/10.18653/V1/2021.FINDINGS-ACL.449.

Feigenblat, G. et al. (2021). TWEETSUMM - A Large Scale Dialog Summarization Dataset for Customer Service. 7 November 2021. Available from https://research.ibm.com/publications/tweetsumm-a-large-scale-dialog-summarization-dataset-for-customer-service [Accessed 7 February 2023].

Feng, Xiachong, Feng, Xiaocheng and Qin, B. (2022a). A Survey on Dialogue Summarization: Recent Advances and New Frontiers. Available from https://pypi.org/project/pyrouge/ [Accessed 29 October 2022].

Feng, Xiachong, Feng, Xiaocheng and Qin, B. (2022b). MSAMSum: Towards Benchmarking Multi-lingual Dialogue Summarization. *DialDoc 2022 - Proceedings of the 2nd DialDoc Workshop on Document-Grounded Dialogue and Conversational Question Answering, Proceedings of the Workshop*, 1–12. Available from https://doi.org/10.18653/V1/2022.DIALDOC-1.1.

Gao, M. and Wan, X. (no date). DialSummEval: Revisiting Summarization Evaluation for Dialogues. 5693–5709.

Labruna, T. and Magnini, B. (2021). From Cambridge to Pisa: A Journey into Cross-Lingual Dialogue Domain Adaptation for Conversational Agents. *CLiC-it*.

Liu, J. et al. (2021). Topic-Aware Contrastive Learning for Abstractive Dialogue Summarization. *Findings of the Association for Computational Linguistics, Findings of ACL: EMNLP 2021*, 1229–1243. Available from https://doi.org/10.18653/V1/2021.FINDINGS-EMNLP.106.

Liu, Z., Shi, K. and Chen, N.F. (2021). Coreference-Aware Dialogue Summarization. *undefined*.

Park, S., Shin, D. and Lee, J. (2022a). Leveraging Non-dialogue Summaries for Dialogue Summarization. Available from https://doi.org/10.48550/arxiv.2210.09474 [Accessed 2 November 2022].

Park, S., Shin, D. and Lee, J. (2022b). Leveraging Non-dialogue Summaries for Dialogue Summarization. *TU*, 1–7.

Wang, B. et al. (2022). A Focused Study on Sequence Length for Dialogue Summarization. Available from https://doi.org/10.48550/arxiv.2209.11910 [Accessed 2 November 2022].

Wang, J. et al. (no date). CLIDSUM: A Benchmark Dataset for Cross-Lingual Dialogue Summarization. Available from https://github.com/krystalan/ [Accessed 18 October 2022].

Xiang, L. et al. (2021). Zero-Shot Deployment for Cross-Lingual Dialogue System. *undefined*, 13029 LNAI, 193–205. Available from https://doi.org/10.1007/978-3-030-88483-3_15.