INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Multilingual Dialogue Summary Generation System for Customer Services

A Project Proposal by

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.

November 2022

Contents

List of Tables

List of Figures

1. Introduction

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.
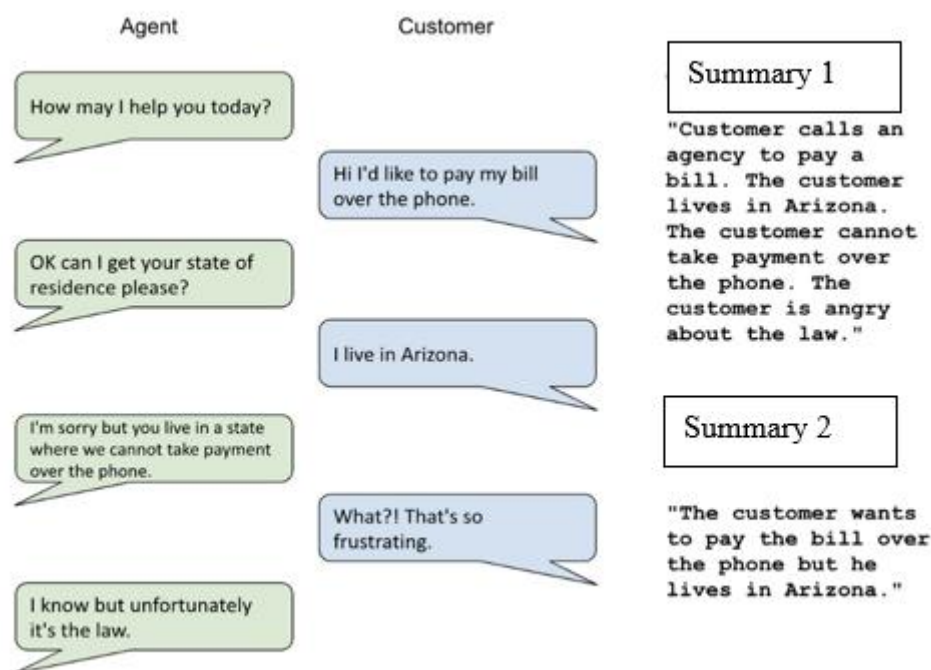
Figure 1 - Dialogue Summary

This proposal aims to provide the reader with an overview 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.

2. Problem domain

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.

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.

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.

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.

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.

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.

4. Research Motivation

The problem identified in this proposal applies mainly to the customer service domain and researchers who have a lot of domain knowledge about Dialogue Summarization. Since dialogue summarization can be utilized in many fields, such as meetings, healthcare, and email threads, this potentially impacts many businesses' use cases. This work is expected to add a valuable contribution to the advancements of the dialogue summarization globally by expanding the possibilities of multilingual Summarization.

5. Existing work

Citation	Brief Description	Limitations	Contribution
(Park, Shin and Lee, 2022)	Transform non-dialogue data into a format exploitable by dialogue summarization models. Due to the lack of diverse datasets in dialogue summarization, they have introduced a method of utilizing the massively available non-dialogue summarization data for enhancing the dialogue summarization systems.	The proposed method can only reinforce the existing dialogue summarization models.	Presents an effective method to utilize abundant nondialogue summarization data to improve dialogue Summarization.

(Zou et al., 2021)	A Domain-agnostic multi-source pre-training paradigm for low resource dialogue summarization.	This proposed method is suitable for dialogue summarization with domain-specific low resources. This has not been considered the low linguistic resources.	This proposed method uses extensive in-domain non-summary data to independently pre-train the dialogue encoder and the summary decoder. To support domain-agnostic summarizing, the combined encoder-decoder model is then pre-trained on the out-of-domain summary data using adversarial criticism. The experimental findings on two open datasets demonstrate that this approach delivers competitive performance and generalizes effectively in various discussion contexts, even with a small amount of training data.
(Liu et al., 2019)	Dialogue summary generation solution for customer service using a Leader-writer network.	Only dialogues with longer key point sequences perform well.	This proposed method uses a novel Leader-writer network with auxiliary key point sequences, ensuring the generated summary is logical and integral.
(Xiang et al., 2021)	Zero-shot Cross-lingual Dialogue Systems with Transferable Latent Variables. They have proposed a transferable latent variable that improves the	The proposed method shows better results in Spanish than Thai. This is because of the quality alignments between the two	The proposed method shows that latent variables cope with the variance of semantically similar sentences across different languages. Cross-lingual transfer between

	zero-short cross-lingual adaption of natural language understanding tasks to low-resource languages.	languages. Spanish is much more lexically and grammatically similar to English than Thai.	English to Spanish and English-to-Thai have demonstrated state-of-the-art results.
(Bai, Gao and Huang, 2021)	Cross-lingual abstractive Summarization with limited parallel resources by sharing summary patterns across languages.	This approach requires a large number of high-quality human-translated parallel summaries.	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.
(Labruna and Magnini, 2021)	Experimental cross-lingual dialogue domain adaptation for conversational agents.	Major performance decrease between English and automatically translated Italian datasets.	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.

Table 1 - Existing work

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.

7. Contribution to the Body of Knowledge

The author's research contribution 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]

7.1 Technological contribution

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.

7.2 Domain contribution

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.

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.

9. Research question/s

RQ1: What are the current limitations in dialogue Summarization?

RQ2: How can existing resources be utilized to develop a multilingual dialogue summarizer using cross-lingual transfer techniques?

RQ3: What are the recent advancements in pre-trained language models that can be used to build a multilingual dialogue summary generation solution?

10. Research Aim

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 elaborate on the research aim, this research project will focus on developing a system that can be utilized using 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.

In order to confirm or refute the selected hypothesis, the necessary knowledge will be examined and researched, components will be developed, and the performance will be evaluated. This multilingual dialogue summary generation platform will be able to run on a hosted server for public use.

11. Research Objective

Research Objectives	Description	Learning Outcomes	Research Questions
Literature Review	Gather required material on related work and critically evaluate the findings. • **RO1:** Study on existing dialogue summarization techniques in the customer service domain. • **RO2:** Study on existing methods of text summarization techniques. • **RO3:** Conduct a preliminary study on pre-trained language models. • **RO4:** Analyze the recent advancements in multilingual models.	LO1, LO4, LO8	**RQ1, RQ2**
Requirement Elicitation	Defining the project's requirements using the proper methods and tools to solve the research problems and gaps should be handled based on relevant prior research knowledge. • **RO1:** Gather information about requirements and resources related to dialogue summarization. • **RO2:** Gather requirements related to the pre-trained language model and understand the capabilities of recent advancements. • **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. • **RO1**: To design a cross-lingual transfer learning model which can be implemented from the existing resources. • **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. • **RO1:** To train the cross-lingual model using the existing resources. • **RO2:** Develop a model to summarize the dialogues. • **RO3:** Develop a user interface and which a user can interact with the application from a browser.	LO1, LO5, LO7, LO8	**RQ2**
Evaluation	Testing the implemented system and dialogue summary generation with evaluation metrics. • **RO1:** Test each component by creating unit tests. • **RO2:** Using performance metrics to evaluate the effectiveness of the summary generation. • **RO3:** To verify the functional and non-functional requirements were made after the research is completed.	LO1, LO5, LO8	**RQ2**

Table 2 - Research Objective

12. Project Scope

The scope of this research project is identified as follows after the study of the previous research work and existing products, considering the period allocated for this project.

12.1 In-scope

The following is the expected scope of this project.

- A graphical user interface in which a user can enter the dialogue in textual format and get the dialogue summary.
- A system that can generate summaries for multiple languages.
- The number of supported languages will be dependent on the selected cross-lingual model.

12.2 Out-scope

The following are the parts that will not be covered in this project,

- Generating summaries for dialogues in audio format.
- Language support for every existing language.
- Ability to integrate this system into an existing platform.

12.3 Prototype Diagram of the proposed system
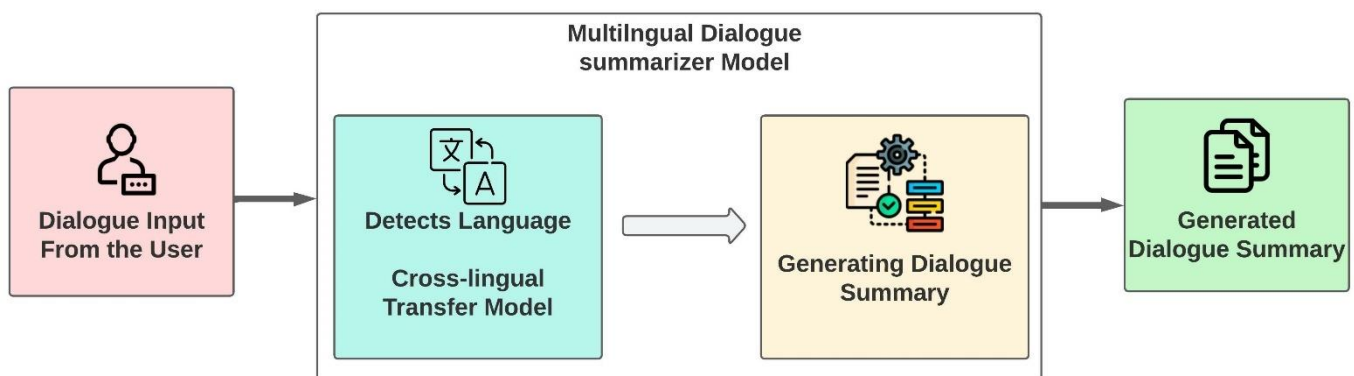
Figure 2 - Prototype Diagram of the proposed system (self-composed)

13. Methodology

13.1 Research Methodology

Research Philosophy	From positivism and interpretivism, the author chooses **pragmatism** as the research philosophy. The selection was determined after considering evaluation methods and previously published research that included quantitative and qualitative data.
Research Approach	This research aims to evaluate and prove the hypothesis. Therefore, the development of cross-language transfer learning for low-resource settings can be a backbone for future research. A **deductive** approach was chosen as the research broadly applies existing theories to the domain of interest.
Research Strategy	The research strategy defines the methodology by answering the research questions. **Surveys** are considered for the research strategy as the primary option. Interviews will also be used as a secondary form of data collection, such as evaluation and feedback, during the final part of the research project.
Research Choice	Among the mono, multi and mixed methods for research choice, the **mixed method was** chosen as both quantitative and qualitative data will be used for this research, such as surveys, performance values and feedback.
Time Horizons	The time horizon determines the duration of the investigation. The **cross-sectional** time

	horizon was chosen from the available two since the data for this study will be collected at a single point in time.

Table 3 - Research Methodology

13.2 Development Methodology

13.2.1 What is the development methodology

A **prototype** model was selected for the research project for the development model. To justify, this research project is developed on a prototype, then the prototype will be tested, and necessary modifications are applied until the desired result is achieved.

13.2.2 Requirement elicitation methodology

Among various requirement elicitation methodologies such as the survey, observation, literature review and brainstorming, for this research project, survey and literature review were selected. Literature reviews and surveys will be used to gather requirements for the development of the prototype.

13.2.3 Design methodology

In this research project, object-oriented analysis and design (OOAD) was selected as the design methodology to reuse the components to extend the system.

13.2.4 Evaluation methodology

Evaluation results are one of the outcomes of research. It can determine the effectiveness of the research findings. Evaluation metrics and benchmarking evaluation approaches will be used to evaluate the multilingual dialogue summarization.

13.2.4.1 Evaluation metrics

The following metrics based on previous research will be used to evaluate the quality of the generated dialogue summaries.

ROUGE(Lin, 2004) – most widely used evaluation metric in Summarization. Further, this can be expanded as ROUGE-1, ROUGE-2, ROUGE-L

BLEU(Papineni et al., 2002) – primary evaluation for machine translations. This can be used to evaluate generated text.

13.2.4.2 Benchmarking

Based on previous research benchmarking for dialogue summarization has been done using many benchmarking datasets. For this multilingual dialogue summarization, ClidSum benchmarking will be used. (Wang et al., 2022)

13.3 Project Management Methodology

Among different project management methodologies, **Prince2** was chosen for this research project as the project scope, and other stages have already been recognized.

13.3.1 Schedule

13.3.2.1 Gantt chart

Figure 3 - Gantt Chart

	Task Name	Start Date	End Date	2023						202...	
				Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
1	⊟ Project Intiation	08/15/22	11/07/22								
2	Selection of research problem and a domain	08/15/22	08/23/22								
3	Supervisor's Initial Idea discussion	08/22/22	08/23/22								
4	Exploring the research domain	08/23/22	08/26/22								
5	Identifying the research gap and the problem domain	08/23/22	09/09/22								
6	Preparing draft project proposal	09/12/22	11/01/22								
7	Supervisors feedback on draft project proposal	10/25/22	11/01/22								
8	Self-evaluation on project proposal	11/01/22	11/02/22								
9	Submission of the project proposal	11/03/22	11/07/22								
10	⊟ Literature Review	08/23/22	11/28/22								
11	Identifying components to cover in the literature review	10/28/22	11/07/22								
12	Collecting related research papers	08/23/22	11/01/22								
13	Reading and summarizing the papers	08/23/22	10/25/22								
14	Prepare Draft LR	11/01/22	11/15/22								
15	Supervisors feedback on draft LR	11/15/22	11/22/22								
16	Finalizing LR based on feedback	11/22/22	11/25/22								
18	⊟ Requirement Gathering	09/12/22	01/02/23								
19	Explore requirement	09/12/22	11/11/22								
20	Create questionnaires	11/16/22	11/18/22								
21	Distribute questionnairs to target audience	11/18/22	11/29/22								
22	Reach to research domain experts for insights	11/14/22	11/29/22								
23	Analyze data from the questionnairs	12/02/22	12/05/22								
24	Preparing SRS	11/11/22	12/23/22								
25	Supervisors feedback on draft SRS	11/22/22	12/27/22								
26	Self evaluation on SRS	12/28/22	12/29/22								
27	Submission of SRS	01/02/23	01/02/23								
28	⊟ Designing the System	11/01/22	02/02/23								
29	Selecting component of the prototype	11/01/22	11/07/22								
30	Creating High-level architecture	11/24/22	12/01/22								
31	Supervisor feedback on High-level architecture	11/29/22	11/29/22								
32	Creating system design	12/05/22	11/11/22								
33	Preparing draft of project specification design (PSD)	11/24/22	01/13/23								
34	Supervised feedback on draft PSD	12/27/22	01/10/23								
35	Self evaluation on PSD	01/13/23	01/13/23								
36	Submission of draft PSD	01/23/23	01/23/23								
37	Submission of PSD	01/30/23	02/02/23								
38	⊟ Prototype Implementation	12/07/22	02/02/23								
39	Define prototype deliverables	12/07/22	12/13/22								
40	Implementing core features	12/19/22	01/10/23								
41	Implementing application components	01/10/23	01/13/23								
42	Demo prototype to Supervisor and get feedback	01/24/23	01/24/23								
43	Refining the prototype	01/25/23	02/01/23								
44	Submission of the Prototype	01/30/23	02/02/23								
45	⊟ Testing and Evaluation	01/12/23	03/17/23								
46	Identifying functional and non-functional requirements	01/12/23	01/17/23								
47	Creating test plans	02/06/23	02/10/23								
48	Writing unit test	02/13/23	02/16/23								
49	Conducting functionality tests	02/17/23	02/21/23								
50	Conducting usability tests	02/22/23	02/24/23								
51	Submission of Test and Evaluation Report	03/20/23	03/20/23								
52	Final evaluation of prototype	02/27/23	03/17/23								
53	⊟ Final Documentation and Conclusion	02/01/23	05/15/23								
54	Drafting the dissertation	02/01/23	04/13/23								
55	Supervisor feedback on draft dissertation	03/14/23	03/28/23								
56	Revisit dissertation	03/28/23	03/29/23								
57	Submission of draft dissertation	03/27/23	03/30/23								
58	Submission of final dissertation	04/24/23	04/27/23								
59	Writing documentation for the product	04/28/23	05/04/23								
60	Project presentation and Viva	05/15/23	05/15/23								

Figure 3 - Gantt Chart

13.3.2.2 Deliverables and dates

Deliverable	Date
Project Proposal Document **The initial proposal of the research project.**	7th November 2022
Ethics Form	7th November 2022
Literature Review Document **A critical review of existing work and solutions.**	27th October 2022
Project Specification Design and Prototype Document **A document specifying the approach of the multilingual dialogue summarizer.**	2nd February 2023
Prototype **A prototype of the research, which includes the multilingual dialogue summarizer**	2nd February 2023
Draft Project Report **A draft version of the final report to get feedback from the supervisor.**	20th March 2023
Final Thesis	27th April 2023

Table 4 - Deliverables and Date

13.3.2 Resource requirements

13.3.2.1 Software resources

- Operating system (Windows/macOS/Linux Distro) – Windows will be the default selection for the development because of overall software availability.

- Base programming language (Python/R) – Python will be the primary programming language for the project because of the availability of open-source packages and libraries.

- System Front-end programming language (Angular/React) – The angular framework will be used to develop the front-end application.

- IDE (Jetbrain PyCharm / Visual Studio) – Jetbrains PyCharm and Visual studio code will be used to support the development environment.

- Cloud-Based Development (Google CoLab/Azure Notebooks) – A cloud-based development environment to develop and train models. Google CoLab will be chosen as it has more RAM by default than the Azure Notebooks.

- Reference Manager (Mendeley/Zotero) – Research management tool to save and back up research resources in a cloud-based environment. Mendeley is chosen as it has both web and desktop versions.

- Cloud Storage to save files (Google Drive/One Drive/Dropbox) – To store necessary files related to this research project. One drive will be the primary selection as it can easily integrate with word applications. Also, google drive will be used as a backup.

- Source Code Management (Github/Gitlab/Bitbucket) – To manage and back up the code related to the research project. GitHub is chosen as it has more available tools than other source code management platforms.

- Continues Integration and Development Services (Amazon AWS, Microsoft Azure) – In order to continuously integrate changes in the implementation of the system a CI/CD pipeline service will required. Amazon AWS is selected as it has more options to choose.

13.3.2.2 Hardware resource

- Core i5 8th Gen processor or above – To perform high CPU resource-intense tasks.
- 50GB Disk space – To store datasets and models for local development.
- 16GB Ram or More – To perform high memory-intensive tasks.

13.3.2.3 Data requirements

- Available datasets on high-resource languages required for the project.

13.3.2.4 Skills requirements

- Skills for understanding mathematical equations.
- Knowledge of natural language processing.
- Knowledge of model evaluation methods.
- Creative report writing skills.

13.3.3 Risk and mitigations

These are the identified risk for this research project with plans for mitigations.

Risk	Probability of Occurrence	The magnitude of the loss	Mitigation Plan
Deep knowledge of language models and cross-lingual transfer learning theories.	5	5	Courses and online resources are available to improve knowledge in these areas.
Updating the resource requirements of the research project	4	2	Using the prototype development approach will be able to identify resource requirements.
Chance of exceeding the free tier in cloud services such as google CoLab and Amazon servers.	3	4	Analyse the available options and select the best plan for the development of the project.
File corruption or losing access to backup files	2	5	Share the backup folder with another personal account in case of losing access to one, so it will be accessible by the backup account.
Inability to conduct the research work due to sickness or illness	1	5	Have a record of project outcomes and maintain good documentation.

Table 5 - Risk and Mitigation

14. References

Bai, Y., Gao, Y. and Huang, H. (2021). Cross-lingual abstractive summarization with limited parallel resources. *ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 6910–6924. Available from https://doi.org/10.18653/V1/2021.ACL-LONG.538 [Accessed 29 October 2022].

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 [Accessed 16 September 2022].

Feigenblat, G. et al. (2021). TWEETSUMM-A Dialog Summarization Dataset for Customer Service. Available from https://github.com/guyfe/Tweetsumm [Accessed 14 September 2022].

Feng, Xiachong, Feng, Xiaocheng and Qin, B. (2021). Incorporating Commonsense Knowledge into Abstractive Dialogue Summarization via Heterogeneous Graph Networks. *undefined*, 12869 LNAI, 127–142. Available from https://doi.org/10.1007/978-3-030-84186-7_9 [Accessed 16 September 2022].

Feng, Xiachong, Feng, Xiaocheng and Qin, B. (2022). 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 [Accessed 17 October 2022].

Feng, Xiachong et al. (2021). Language model as an annotator: Exploring DialoGPT for dialogue summarization. *ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 1479–1491. Available from https://doi.org/10.18653/V1/2021.ACL-LONG.117 [Accessed 16 September 2022].

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].

Gao, M. and Wan, X. (2022). DialSummEval: Revisiting Summarization Evaluation for Dialogues. Available from https://github.com/m-popovic/chrF [Accessed 29 October 2022].

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

Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries.

Liu, C. et al. (2019). Automatic Dialogue Summary Generation for Customer Service. Available from https://doi.org/10.1145/3292500.3330683 [Accessed 17 October 2022].

Papineni, K. et al. (2002). Bleu: a Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, 311–318. Available from https://doi.org/10.3115/1073083.1073135 [Accessed 30 October 2022].

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

Tuggener, D. et al. (2021). Are We Summarizing the Right Way? A Survey of Dialogue Summarization Data Sets. 107–118. Available from https://elitr.github.io/ [Accessed 16 September 2022].

Wang, J. et al. (2022). ClidSum: A Benchmark Dataset for Cross-Lingual Dialogue Summarization. Available from https://doi.org/10.48550/arxiv.2202.05599 [Accessed 30 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 [Accessed 29 October 2022].

Zhang, Y. et al. (2021). An Exploratory Study on Long Dialogue Summarization: What Works and What's Next. *Findings of the Association for Computational Linguistics, Findings of ACL: EMNLP 2021*, 4426–4433. Available from https://doi.org/10.18653/V1/2021.FINDINGS-EMNLP.377 [Accessed 16 September 2022].

Zou, Y. et al. (2021). Low-Resource Dialogue Summarization with Domain-Agnostic Multi-Source Pretraining. Available from http://arxiv.org/abs/2109.04080 [Accessed 18 October 2022].