DESIGN AND DEVELOPMENT OF A SINHALA BANKING CHATBOT SYSTEM

P.L.A.S. Maleesha

2024

Design and Development of a Sinhala Banking Chatbot System

P.L.A.S. Maleesha
2024

Design and Development of a Sinhala Language Banking Chatbot System

A dissertation submitted for the Degree of Master of Computer Science

P.L.A.S. Maleesha

University of Colombo School of Computing

2024

Declaration

Name of the student: P.L.A.S.Maleesha
Registration number: 2020/MCS/053
Name of the Degree Programme: Master of Computer Science
Project/Thesis title: Design and Development of a Sinhala Language Banking Chatbot System

1. The project/thesis is my original work and has not been submitted previously for a degree at this or any other University/Institute. To the best of my knowledge, it does not contain any material published or written by another person, except as acknowledged in the text.

2. I understand what plagiarism is, the various types of plagiarism, how to avoid it, what my resources are, who can help me if I am unsure about a research or plagiarism issue, as well as what the consequences are at University of Colombo School of Computing (UCSC) for plagiarism.

3. I understand that ignorance is not an excuse for plagiarism and that I am responsible for clarifying, asking questions and utilizing all available resources in order to educate myself and prevent myself from plagiarizing.

4. I am also aware of the dangers of using online plagiarism checkers and sites that offer essays for sale. I understand that if I use these resources, I am solely responsible for the consequences of my actions.

5. I assure that any work I submit with my name on it will reflect my own ideas and effort.I will properly cite all material that is not my own.

6. I understand that there is no acceptable excuse for committing plagiarism and that doing so is a violation of the Student Code of Conduct.

Signature of the Student	**Date (DD/MM/YYYY)**
	19/09/2024

Certified by Supervisor(s)

This is to certify that this project/thesis is based on the work of the above-mentioned student under my/our supervision. The thesis has been prepared according to the format stipulated and is of an acceptable standard.

	Supervisor 1	**Supervisor 2**	**Supervisor 3**
Name	Dr. B. H. R. PUSHPANANDA		
Signature			
Date	23 - 09 - 2024		

I dedicate this thesis to my wonderful supervisor, whose wise counsel, deep knowledge, and steadfast support have been essential to this project's success. In addition to inspiring my research journey, his creative vision and unwavering commitment to academic achievement have also shaped my professional ethos. This work is evidence of his tremendous influence on my academic and personal development, as well as his priceless mentoring. I am incredibly grateful for all he has done for me academically and professionally. He is the epitome of what it means to be a great mentor.

ACKNOWLEDGEMENT

I am really grateful to all those who have been instrumental in the accomplishment of this thesis, which is titled “Design and Development of a Sinhala Language Banking Chatbot System”. Special thanks go out to my supervisor Dr. Randil Pushpananda, whose wise counsel, unflinching tolerance, and perceptive criticism have served as the foundation for this study. My sincere gratitude is extended to my colleagues, whose helpful conversations and collaborative attitude have greatly enhanced the scope and depth of this research.

Moreover, I am immensely grateful to my family for their endless love, support, and encouragement. Their belief in my capabilities has been a constant source of motivation, pushing me to excel even in the face of adversity.

This thesis represents not just my efforts, but the collective support, wisdom, and encouragement of all the individuals mentioned above. To them, I owe a debt of gratitude that this acknowledgment can only begin to express.

ABSTRACT

The language inclusion of banking services is still a crucial concern in an era where digital banking is becoming more and more common, especially for people who do not understand English. In order to bridge the linguistic gap in Sri Lanka's banking industry, this thesis presents a novel banking chatbot system in the Sinhala language. Through the utilization of OpenAI's Large Language Models, cutting-edge translation APIs, and the Rasa framework, this research leads the way in creating a chatbot that can comprehend and reply to user inquiries in Sinhala, thereby greatly improving the accessibility of banking services for Sinhala speakers.

The comprehensive integration of natural language processing techniques used in this research's approach entails translating Sinhala questions into English, processing those queries to provide pertinent responses, and then translating those responses back into Sinhala. This method preserves the linguistic and cultural subtleties of the user's original query while simultaneously guaranteeing the correctness and applicability of the chatbot's responses. The architecture of the chatbot is made to be reliable, scalable, and able to handle a variety of banking-related queries, such as those for transactional services or account information.

Keywords: Sinhala Language, Digital Banking, Chatbot Technology, Rasa Framework, OpenAI, Large Language Models, Natural Language Processing, Translation API, User Experience, Linguistic Inclusivity.

ABBREVIATIONS

AI	-	Artificial Intelligence
API	-	Application Programming Interface
Docker	-	A platform for developing, shipping, and running applications
GCP	-	Google Cloud Platform
IoT	-	Internet of Things
JS	-	JavaScript
LLM	-	Large Language Models
Nginx	-	Engine X (a web server)
NLP	-	Natural Language Processing
NLU	-	Natural Language Understanding
OpenAI	-	An AI research and deployment company
Rasa	-	An open-source machine learning framework for automated text and voice-based conversations
UI	-	User Interface

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The integration of artificial intelligence in customer service, particularly via chatbots, is a noteworthy advancement in automating and improving consumer interactions across multiple industries, including banking. But because technology is so universal, it frequently ignores the linguistic variety and unique requirements of local communities, leading to a digital divide. The lack of locally tailored digital banking options prevents Sri Lanka, where Sinhala is the primary language spoken, from fully utilizing these technical breakthroughs. In order to bridge the gap between financial inclusion and linguistic technology, this thesis developed a banking chatbot system in Sinhala.

The effort to incorporate Sinhala into a banking chatbot framework is more than just a technological challenge; it's a step toward democratizing financial services accessibility and guaranteeing that language hurdles won't stop people from becoming digitally empowered. Through the use of the Rasa framework and OpenAI's Large Language Models, which are translated via complex APIs, this project aims to provide a user experience that is both smooth and easy to use for the Sinhala-speaking community (McTear and Ashurkina, 2024). This breakthrough is significant because it promotes linguistic diversity and inclusion by improving the user experience and acting as a model for incorporating other minority languages into digital services.

Furthermore, by describing the process of customizing cutting-edge technology to satisfy particular linguistic and cultural needs, this thesis adds to the body of knowledge on computational linguistics, AI and user experience design in academic circles (Dwivedi *et al.*, 2023). The study offers empirical support for the viability and significance of localized AI-driven banking solutions through a thorough assessment of the chatbot's performance, user satisfaction, and language accuracy.

This study advocates for a more inclusive approach to technology development and seeks to inspire future discoveries in the field, while also tackling a major technological and societal concern. This thesis emphasizes the significance of linguistic inclusion in the digital era and the potential of AI and NLP technologies to overcome language barriers in order to promote a more accessible and equitable digital environment.

1.1 Motivation

The motivation for this thesis is deeply intertwined with the technological evolution within the banking sector, emphasizing the critical role of chatbots in reshaping customer interactions. The use of technology like Internet banking, AI, and the IoT to modernize banking globally, and particularly in Sri Lanka, is a significant step towards providing more effective and accessible customer service. The majority of English in these systems, however, poses a serious obstacle for non-English speakers, particularly in Sri Lanka's Sinhala-speaking populace. This divide makes it difficult to communicate effectively and restricts access to cutting-edge banking services, which forces people to rely on older, less effective banking practices.

In response, the thesis proposes to break down linguistic barriers by being the first to design a banking chatbot that speaks Sinhala. This project aims to raise user satisfaction, promote financial inclusiveness, and bring Sri Lanka into line with international banking technology breakthroughs. The potential to provide smooth banking interactions in Sinhala is what drives the research. This would guarantee that all people have equal access to digital banking advances and further the larger objective of improving the accessibility and usability of technology-driven financial services. This research pushes the envelope of technological inclusivity in the financial industry by addressing a pressing language need and laying the groundwork for future advancements in multilingual chatbot systems.

1.2 Statement of the Problem

The fundamental lack of Sinhala LLMs appropriate for direct integration with banking chatbots is the research's stated difficulty. This disparity poses a serious obstacle to the creation of digital banking systems that can serve consumers who speak Sinhala and are both linguistically inclusive and responsive. In addition to impeding the development of effective and user-friendly chatbot services, the dearth of Sinhala LLMs widens the digital divide by depriving a significant segment of Sri Lankan society of sophisticated banking technology that meet their linguistic requirements. Solving this problem is essential to promoting financial inclusion and guaranteeing that technical developments in banking are equally available to all language groups in the nation.

1.3 Research Aim and Objectives

With the goal of completely changing the digital banking experience for Sri Lankans who speak Sinhala, the research aims to be a trailblazer in the integration of Sinhala language capabilities into banking chatbots. The purpose of this research is to use advanced language processing technology to design, create, and assess a banking chatbot in Sinhala. This project aims to ensure inclusivity and improve user engagement across linguistic differences by democratizing access to digital financial services by tackling the large language barrier seen in current banking technologies.

1.3.1 Aim

The aim of this project is to construct a banking chatbot system that can comprehend and answer in Sinhala, so offering a transformational solution to the Sinhala-speaking community. This project aims to utilize state-of-the-art technology in AI and NLP, with a specific focus on addressing the present obstacles caused by the lack of Sinhala-specific LLMs. By doing this, the project hopes to improve the inclusiveness and equity of digital banking services and make them more accessible, allowing Sinhala speakers to take full advantage of and participate in technology improvements in the banking industry. This project not only closes a big technology divide but also advances the more general objectives of language diversity and financial inclusiveness in digital networks.

1.3.2 Objectives

The project aims to systematically address the obstacles and take advantage of the opportunities in the process of creating a banking chatbot system in the Sinhala language. These goals stem from a thorough grasp of the state of technology, the unique requirements of the Sinhala-speaking community, and the language obstacles that currently exist in digital banking services. This research aims to make a significant impact in the field of digital banking by utilizing a strategic approach that includes design, development, integration, evaluation, and contribution to technological inclusivity. This approach ensures that the advantages of technological advancements are available to everyone, irrespective of language. In order to provide a comprehensive and significant study, this preparatory work establishes the foundation for accomplishing the specific objectives, each of which aims to address a certain aspect of the research aim. Therefore, the objectives identified for this project are as follows,

- **Design & Development :-** To build an intelligent chatbot that can comprehend and respond to Sinhala queries using the Rasa framework. In order to ensure accurate and contextually appropriate interactions, this entails integrating translation APIs with OpenAI's LLMs to translate user inquiries from Sinhala to English, process them, and then translate the responses back to Sinhala.

- **Linguistic Integration :-** To address the issue of the lack of Sinhala LLMs by coming up with creative ways to enable the chatbot to effectively support the Sinhala language. This entails investigating and creating techniques to precisely decipher and produce Sinhala language answers, therefore getting over the limitations of the available technology.

- **Performance Evaluation :-** To carry out thorough analyses of the chatbot's operation, evaluating its precision, speed of response, and applicability of answers to user inquiries in Sinhala. Surveys and interviews with users will be used to get their input in order to gauge their level of satisfaction and pinpoint areas that need work.

- **Accessibility Improvement :-** By making the chatbot user-friendly and simple to use, we hope to improve the accessibility of digital financial services for the Sinhala-speaking community. In order to promote greater financial inclusion, this target focuses on removing language barriers that have traditionally restricted access to digital banking.

- **Contribution to Technological inclusiveness :-** To establish a standard for upcoming advancements in multilingual chatbot systems and to advance the field of linguistic inclusiveness in technology. In order to spur more advancements in this field, this initiative intends to show the viability and advantages of including underrepresented languages into digital services.

1.4 Scope

The scope of the research includes designing and developing a text-based AI chatbot specifically for Sinhala-speaking bank customers. The chatbot uses cutting-edge natural language processing algorithms for entity recognition and purpose detection in order to comprehend and react to questions in Sinhala. Utilizing translation APIs to translate Sinhala input into English and integrating the Rasa framework for dialogue management, the chatbot makes use of OpenAI's advanced models to produce pertinent responses. After that, the answers are translated back into Sinhala, allowing for a smooth communication that improves the user experience. In order to facilitate simple interactions between the user and the chatbot and make the system accessible and understandable for all users, a great deal of attention is spent on creating an intuitive web interface. Exclusions from the scope are explicitly noted to maintain a focused effort on the project's core objectives. This project does not cover voice-based interactions or sophisticated natural language processing beyond the stated intent and entity recognition. This boundary guarantees that the project stays within realistic parameters and focuses on improving text-based communication, which is essential for the chatbot system's successful acceptance and efficacy. By laying out these requirements precisely, the project is positioned to focus on the most important areas and provide a solid solution that helps Sinhala speakers bridge the language gap in digital banking.

1.5 Structure of the Thesis

Chapter 2 **-** This chapter presents a comprehensive review of the literature on integrating Sinhala language capabilities into chatbots, focusing on the development of techniques within the banking sector.

Chapter 3 **-** Describes the process of developing a Sinhala Language Banking Chatbot System using advanced NLP frameworks like Rasa and integrating translation services and OpenAI's LLMs. It discusses preprocessing methods such as translation and intent recognition, and the system architecture that facilitates seamless interaction between users and the machine learning model through a user-friendly web interface.

Chapter 4 **-** Evaluates the effectiveness and efficiency of the Sinhala Language Banking Chatbot System, emphasizing the importance of accurate and user-friendly chatbot responses in the context of increasing digital banking usage among Sinhala-speaking users.

CHAPTER 2
LITERATURE REVIEW

The "Design and Development of a Sinhala Language Banking Chatbot System" thesis's literature review chapter delves deeply into the field and critically examines previous research on chatbots, NLP, AI in banking, and the technological integration of underrepresented languages like Sinhala. This chapter tries to outline the development of AI and machine learning in the banking sector, stress the need of creating multilingual chatbots, and point out the research gaps that your study fills. It explores a range of techniques and tools, including the Rasa framework and LangChain with RAG models for Sinhala language processing, highlighting the creative way in which your study addresses linguistic constraints in digital financial services. By placing the study within the larger academic discourse, this review not only lays the groundwork for the work but also highlights how it has advanced the subject of linguistically inclusive banking solutions.

2.1 Literature Review

By presenting the first-ever Sinhala chatbot system, the study by Hettige and Karunananda represented a noteworthy turning point in the fields of artificial intelligence and natural language processing(Hettige and Karunananda, 2006). This innovative project filled a significant void in chat system technology by providing a solution designed with the Sinhala-speaking community in mind. In order to create a human-computer dialog system that can comprehend and reply to inquiries in Sinhala, the research expanded the capabilities of an already-existing Sinhala parsing system, which was originally created for machine translation from English to Sinhala. This invention highlighted the significance of removing language barriers in technology and increased the accessibility of digital communication for non-English users, while also showcasing the adaptability of the underlying parsing mechanism. The study did, however, also point out certain shortcomings in the Sinhala Chatbot's current implementation. The chatbot was primarily limited to answering general questions regarding date, time, and user identity because its knowledge base contained only basic operating system information. Taking note of this, the researchers determined that additional work was required to increase the chatbot's knowledge base and broaden its scope of use. This indicates that much more study is needed to improve the chatbot's ability to respond to a wider variety of customer inquiries, particularly in specialized industries like financial services. Future research must focus on expanding the chatbot's use to domain-specific applications, which calls for the creation of a more thorough and specialized knowledge base to guarantee accuracy and relevance in particular situations. Furthermore, the study's recommendation to investigate the server-side implementation of the Sinhala chatbot presents an

additional line of inquiry. This concept emphasizes how server-side architecture may help chatbot systems become more scalable, effective, and accessible. An architectural approach like this might have a big impact on chatbot deployment and maintenance, providing a more reliable and adaptable way to meet different user needs. These highlighted research gaps and limitations offer significant chances for additional investigation and improvement as the fields of artificial intelligence and natural language processing continue to develop, offering potential breakthroughs in the development of more intelligent, responsive, and user-centered dialog systems.

The creation of a Smart Chatbot System in the research by Amama and Okengwu signifies a noteworthy breakthrough in the field of digital banking(Amama and Okengwu, 2023). The promise of artificial intelligence to improve customer service in the banking industry is demonstrated by this system, which combines machine learning (ML) with natural language processing (NLP) algorithms. The research establishes a new standard for technical innovation in online banking services by demonstrating the accuracy and dependability of the chatbot in interpreting and responding to user queries, with a high F1 score of 0.97. An example of the transformative power of NLP and ML technologies in streamlining customer interactions and administrative processes is the creation of an intuitive interface, which enhances accessibility and efficiency of banking operations. But the study also points out particular shortcomings and areas where the current system needs to be improved. These include adding a logging system for administrative control, improving the chatbot's correctness, and integrating system memory and anaphora resolution for improved context understanding. These acknowledgements draw attention to the intricate difficulties involved in developing intelligent chatbots that can adjust to changing user needs and keep up with technology breakthroughs. These drawbacks imply that ongoing research and development are necessary to hone and enhance the functioning of chatbots used in the banking sector. The report also identifies important research gaps that need to be investigated further. These include looking at anomalies that could have an impact on classification accuracy, improving the chatbot's capacity to resolve references and maintain context during an exchange, and emphasizing the value of a strong logging system for analytics and monitoring. Closing these gaps may result in chatbots that are more intelligent, dependable, and focused on the needs of the user. This study supports my own proposal to create a text-based AI chatbot for Sinhala-language banking services. It also shows how NLP and ML technologies can be integrated to revolutionize the banking industry and make digital banking more accessible and effective.

The study by Wicaksono and Zahra offers a critical examination of the adoption of chatbot technology within the Indonesian banking sector, focusing on navigating the challenges of regulatory compliance, language preferences, and industry-specific requirements(Zahra, 2022). By evaluating two prominent chatbot platforms, Rasa and Botika, the research contributes valuable insights into selecting an appropriate platform based on usability and Human-Machine Symbiosis and Agent Model (HMSAM) evaluations. The recommendation of Botika, in particular, underscores the study's practical implications for Indonesian banks considering the deployment of chatbots to streamline customer service operations and potentially reduce operational costs. This research is pivotal in highlighting the potential for chatbots to enhance or even replace traditional call center services within the banking industry, pointing towards a more efficient and cost-effective approach to customer service. The research acknowledges limitations despite its contributions, mostly related to the narrow scope of its evaluation, which is limited to two chatbot platforms. This limitation points to a larger research opportunity to investigate other platforms that might provide better compatibility with the unique requirements of Indonesian banks, such as those that better handle language preferences and regulatory limits. The admission of these drawbacks emphasizes how difficult it is to integrate chatbot technology into the banking industry, pointing to the necessity for a more thorough examination of the platforms that are currently available and their individual capacities in order to satisfy the complex requirements of Indonesia's banking environment. The report also cites a number of research gaps, such as the requirement for a more thorough assessment of chatbot platforms, an in-depth investigation of implementation difficulties, and an evaluation of the long-term effects of chatbot technology on the banking industry. By filling in these gaps, we can ensure that banks can choose and deploy chatbot solutions that not only adhere to regulatory requirements and cultural preferences but also improve long-term operational efficiency and customer satisfaction. This will advance the understanding and application of chatbot technology in banking. Thus, Wicaksono and Zahra's research (2022) lays the groundwork for future research into these crucial areas and contributes to the wider use and optimization of chatbot technology in the Indonesian banking industry.

The study by Dissanayake et al. (2022) introduces Kolloqe, a pioneering no-code chatbot development platform tailored for the Sinhala-English bilingual speakers, a significant step towards addressing the nuanced challenges of code-switching and code-mixing in natural language processing(Dissanayake *et al.*, 2022). Kolloqe stands out by offering a suite of built-in tools designed for natural language processing, model evaluation, and notably, explainability - a feature often neglected in current chatbot development platforms. This platform not only facilitates the development of more effective and relevant chatbots for the Sinhala-speaking

community but also advances the field by emphasizing the importance of model transparency and interpretability. The study does acknowledge certain potential restrictions on Kolloqe's scalability and its use in a variety of real-world contexts. The ability of the platform to manage intricate and dynamic conversational situations warrants additional investigation, indicating that although Kolloqe represents a noteworthy development, there is still opportunity for enhancement to fully actualize its possibilities in wider uses. These drawbacks draw attention to the continued difficulties in developing flexible and reliable chatbot platforms that may meet the complex needs of multilingual or bilingual communities. The research gaps identified, including scalability, real-world deployment, adaptability to dynamic conversational contexts, and user experience, outline a roadmap for future investigations. These areas underscore the necessity of enhancing Kolloqe to meet the evolving demands of chatbot technology, particularly in ensuring that it can effectively manage a large volume of users and a wide array of conversational nuances. Furthermore, a deeper dive into the user experience and interaction with Kolloqe-developed chatbots could offer valuable insights into its comparative advantages and areas for refinement. As such, the contributions of Kolloqe lay a solid foundation for advancing chatbot development for low-resource languages, paving the way for more inclusive and accessible conversational AI technologies.

In conclusion, while the research contributes significantly to the understanding of adopting Chatbot technology in Indonesian banks, there are research gaps in conducting a more comprehensive platform evaluation, exploring detailed implementation challenges, and assessing the long-term impact of Chatbot implementation. Addressing these gaps will enhance the effectiveness and sustainability of Chatbot technology in the Indonesian banking industry.

2.2 Large Language Models (LLM)

Recent developments in the quickly developing fields of artificial intelligence and natural language processing have been largely dependent on Large Language Models (LLMs)(Leocadio, 2023). These models are excellent at a wide range of tasks, from text production and language translation to conversational AI and content creation(Raiaan *et al.*, 2023). They are renowned for their massive neural networks and thorough training on enormous datasets. Specifically trained on mostly English data to support a broad range of applications targeted at English-speaking customers, English LLMs have experienced substantial development.

However, it's important to note that as of now, there has been no implementation of a Large Language Model specifically for the Sinhala language. Millions of people in Sri Lanka speak Sinhala, which poses special difficulties for natural language processing because of its intricate grammatical structure, use of script, and comparatively small corpus of digital language resources

when compared to English. A gap in language technology development is brought to light by the lack of a Sinhala LLM, emphasizing the necessity of more inclusive research and development initiatives that cover a wider spectrum of languages.

Sinhala speakers do not yet have access to the same degree of advanced AI-powered products and applications as English speakers do due to the lack of a Sinhala-specific LLM. This discrepancy highlights the larger problem of linguistic inequality in AI research, as low-resource languages are frequently marginalized. It will need committed work to gather Sinhala language datasets, create Sinhala language models, and eventually deploy LLMs that can assist the Sinhala-speaking community as effectively as English LLMs in order to address this discrepancy. Such advancements would support accessibility and linguistic diversity in technology in addition to advancing the science of artificial intelligence.

2.3 Chatbot with LLM

Integrating Large Language Models (LLMs) with chatbot technology represents a significant advancement in the field of conversational AI (Govindasamy, 2023). LLMs, with their vast neural networks trained on extensive datasets, possess the ability to understand, generate, and interact with human language in a remarkably sophisticated manner. This capability makes them particularly well-suited for enhancing chatbot applications, enabling these systems to conduct more natural, contextually aware, and nuanced conversations with users.

AI and NLP have advanced significantly with the addition of LLMs to chatbot technology(Belda-Medina and Calvo-Ferrer, 2022). These advanced AI models, which are based on large neural networks and have been trained on enormous datasets, have significantly improved the capabilities of chatbots, making them more responsive, intuitive, and effective at managing human language. With the use of LLMs, chatbots are able to carry out a vast range of language-related tasks, including comprehending context, producing responses that seem human, translating between languages, summarizing content, and engaging in conversational AI at a degree of complexity and nuance that was before unachievable. As a result, chatbots have evolved from basic rule-based responders to sophisticated virtual assistants that can offer a variety of services like educational materials, personal help, and customer support in a way that resembles communication with a human. As a result of LLMs' constant evolution and improvement, chatbot apps continue to advance, becoming more complex instruments with exceptional accuracy for understanding and producing English text. This development demonstrates how LLMs have the power to completely transform the way we engage with digital systems by providing consumers in a variety of industries with tailored and contextually aware experiences.

2.4 Banking chatbot with LLM

LLM enhanced banking chatbots represent a quantum leap in the integration of artificial intelligence with financial services, providing a degree of engagement that nearly emulates human speech(Maseke, 2024). Through the use of sophisticated neural networks and large datasets for training, LLMs' broad capabilities enable these chatbots to understand, interpret, and react to natural language queries with previously unheard-of levels of accuracy(Aslam, 2023). Thanks to technological advancements, chatbots may now perform a wide range of banking functions, such as account administration, transaction queries, financial advising, and seamless client support. The consumer banking experience is revolutionized by these chatbots' capacity to comprehend complex customer requests and deliver prompt, pertinent responses, increasing its efficiency and accessibility 24/7.

Additionally, the addition of LLMs to banking chatbots improves customer experience while also bringing additional operational security and efficiency for financial institutions. Chatbots that use are capable of handling confidential data with care, abiding by tight financial guidelines, and protecting user privacy. They continuously learn from their interactions with consumers and make improvements, so they may eventually provide more individualized services. This lowers operational costs by eliminating the need for large human customer care teams and greatly improves client happiness by offering quick, accurate, and highly customized banking solutions. Banking chatbots are set to become even more essential to the digital banking environment as LLM technology develops, completely changing how consumers get and interact with financial services.

2.5 The Need for Sinhala Language Banking Chatbots

The necessity to address linguistic diversity and accessibility challenges encountered by Sri Lanka's Sinhala-speaking population has led to a need for banking chatbots in the language. The development of Sinhala language chatbots is essential because Sinhala is one of the official languages of the nation and a sizable fraction of the populace does not speak English, which is the language used by the majority of banking chatbots now in use. Financial services delivered by chatbots in Sinhala would greatly improve linguistic accessibility, guaranteeing that a larger population—including non-native English speakers could effectively obtain banking services.

This move towards linguistic inclusivity is crucial in a multiethnic and multilingual nation like Sri Lanka, where catering to the linguistic preferences and cultural sensitivities of diverse language groups is essential for fostering inclusivity and diversity within the banking sector.

Moreover, providing banking services in Sinhala can significantly enhance user satisfaction by streamlining financial interactions and transactions and making them more comprehensible for Sinhala speakers. By removing linguistic barriers, this improvement in communication can result in increased customer satisfaction and engagement. Furthermore, by offering information on financial products and services, individualized financial advice, and support in making informed financial decisions, Sinhala banking chatbots have the potential to enhance financial literacy among the Sinhala-speaking population, thereby fostering financial inclusion and empowerment. Additionally, banks can get a competitive advantage by setting themselves apart from rivals and drawing in more business from customers by implementing chatbots in the Sinhala language. By prioritizing language inclusion and customer-centricity, banks can strengthen their relationship with Sinhala-speaking customers and foster brand loyalty, underscoring the critical need for the development of banking chatbots in the Sinhala language.

CHAPTER 3
METHODOLOGY

The methodological section describes the methodical process used to create a banking chatbot system in Sinhala. The processes involved in choosing tools and technologies, gathering and preparing data, creating the system architecture, creating the user interface, and launching the chatbot are succinctly summarized in this section. This research guarantees the efficient deployment of the chatbot system and its integration with Sinhala Language Models (LLMs) to provide users with correct and contextually relevant responses by adhering to a defined methodology.

3.1 Selection of Tools and Technologies

An essential part of the process for creating a banking chatbot in Sinhala is the "Selection of Tools and Technologies" stage. During this stage, the technologies, frameworks, and tools required to properly develop, implement, and launch the chatbot must be carefully considered and selected.

Choosing the appropriate framework is essential for developing a chatbot, especially one that addresses the linguistic quirks of Sinhala speakers. Rasa's open-source design, strong dialogue management support, and NLU make it an excellent option(*Rasa: Developer Documentation Portal*, no date). Because of its scalability and adaptability, it is the best framework available for creating conversational AI applications(Permadi, 2023). Its extensive ecosystem also helps with the development process. Using pre-existing language models such as OpenAI's Rag becomes essential when tackling the integration difficulty of Sinhala, a language for which specialized LLMs are not easily accessible. This method ensures contextually relevant responses even in the face of linguistic barriers by leveraging the model's comprehension and fine-tuning based on prompts pertaining to banking to handle queries on specific financial issues.

In addition, the lack of direct Sinhala LLMs makes it necessary to employ translation APIs, such Google Translate, in order to overcome the language barrier(Wade, 2011). In order to provide a smooth user experience, this technique entails translating user queries from Sinhala to English, processing them through the language model, and then translating the responses back to Sinhala.

The choice of libraries and development tools is crucial to the operationalization of the chatbot system. For example, using Python for back-end work and HTML, CSS, and JavaScript for front-end development boosts the functionality of the selected frameworks and benefits the chatbot. JQuery and other libraries may greatly improve UI interaction and improve the user experience.

Lastly, for deployment, platforms like the Google Cloud Platform (GCP) offer robust infrastructure, while containerization techniques such as Docker ensure the application's portability and scalability. The use of a web server like Nginx for UI serving further guarantees an efficient and responsive interface, making these technological choices not just practical but essential for the development of a high-performing, Sinhala-compatible banking chatbot.

3.2 Data Collection and Preprocessing

Preprocessing and data gathering are crucial phases in the development of a chatbot system because they guarantee that the system has access to appropriately labeled and pertinent data for testing and training. This explains the steps involved in gathering and preparing data, taking into account both the chatbot's original design and OpenAI integration.

3.2.1 Data Collection for Initial Chatbot Development

When developing a chatbot, especially one tailored for the Sinhala-speaking demographic, the initial step involves the meticulous collection of a dataset that encompasses a broad spectrum of banking-related questions and answers. This crucial dataset forms the foundation for training the chatbot's NLU component, equipping it with the capability to accurately discern user intents and extrapolate pertinent information from inquiries(Casanueva *et al.*, 2022). The process of assembling this dataset often taps into a variety of sources, such as historical customer support interactions, frequently asked questions (FAQs), banking documentation, and discussions on online forums. This diversified approach to data collection is essential to encapsulate the extensive range of topics and user queries related to banking, ensuring the chatbot can handle a wide array of customer interactions. Following data acquisition, a series of preprocessing activities, such as data cleaning, normalization, and annotation becomes imperative. These steps are undertaken to enhance the quality of the dataset, eliminating any errors or inconsistencies, and meticulously labeling the data with appropriate intents and entities. Such preparatory measures are pivotal in training a robust and effective chatbot model, designed to meet the specific needs of users engaging with banking services in Sinhala.

3.2.2 Data Collection for Testing after Integration of OpenAI

When OpenAI's Rag model is integrated with the chatbot to generate responses, it becomes much less dependent on a large dataset because the model can create responses based on prompts instead of a pre-compiled training set (Yusuf, Karaarslan and Aydın, 2024). However, it is still required to keep a small collection of test data in order to assess the system's overall performance and adjust its parameters as needed. This test data, which includes a range of example user queries and the expected responses to them, covers a number of banking scenarios and user intents. It is an essential tool for evaluating the coherence, correctness, and relevance of the chatbot's responses when interacting with users. This test data is subjected to extensive preprocessing to guarantee its consistency, accuracy, and applicability, much like in the first stages of data gathering. These steps include verifying the expected response correctness, refining prompts for optimal response generation with the Rag model, and rigorously testing the system's performance across different scenarios, thereby ensuring the integrated chatbot system operates effectively and efficiently in real-world banking applications.

3.3 Implementation of Baseline System with Rasa

The implementation of a baseline system with Rasa involves setting up and configuring the Rasa framework to create a basic chatbot that can understand user intents and respond accordingly.

To develop a chatbot using Rasa Open Source, begin by adhering to the official installation instructions, typically involving the use of pip to install necessary programs. Subsequently, employ the *rasa init* command to generate a new Rasa project directory, which includes the essential boilerplate code and file structure. Within this framework, the chatbot's domain is defined, encompassing slots, actions, responses, intents, and entities where entities represent pieces of information extracted from user messages, and intents signify the underlying goals or intentions. Slots are utilized to store crucial data derived from user interactions, while actions are defined as the chatbot's possible responses to user messages. The next step involves preparing training data, including narratives and NLU data; narratives illustrate hypothetical dialogues to outline interaction flows, and NLU data comprises user messages annotated with intents and entities for training the NLU model. Following this preparation, use the rasa train NLU command to train the NLU model, enabling it to categorize user messages and extract entities. Further, the dialogue management model is trained using the rasa train command, which crafts a machine learning model designed to predict the chatbot's next actions based on conversation context. To assess the chatbot's performance, interact with it via the command-line interface or Rasa shell, focusing on its ability to understand user intents, extract entities, and generate appropriate

responses. This testing phase is crucial for identifying and implementing necessary adjustments to the model configurations and training data, aiming to enhance the chatbot's accuracy and efficiency. The model summary is denoted in below Figure 3-1.

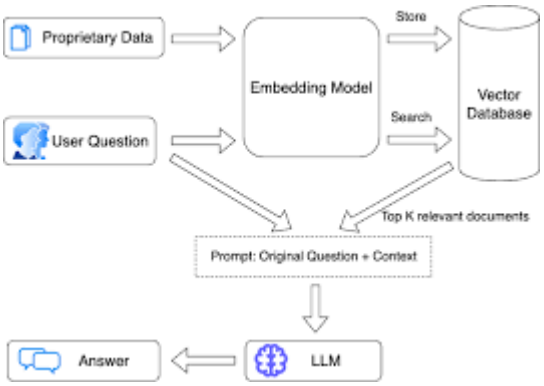


Figure 3-1 Rasa Chatbot Structure

3.4 System Architecture Design

The primary goal of this research is to create a chatbot system tailored specifically for the Sinhala-speaking population, addressing the linguistic barriers and enhancing accessibility to banking services. Initially, the project utilized the Rasa framework, where questions and answers were manually added to provide responses to user queries. However, recognizing the potential of Sinhala Language Models (LLMs) in improving accuracy and ease of interaction, the project evolved to integrate LLMs into the chatbot UI.

Since dedicated Sinhala LLMs were unavailable, the project adopted a unique approach. User questions were translated from Sinhala to English using the Google Translate API before being passed to OpenAI's Rag model. This allowed the system to generate contextually relevant responses related to specific banking topics. To ensure a seamless user experience, the responses from OpenAI were then translated back into Sinhala before being presented to the user within the chatbot UI. The integration of OpenAI's Rag model was facilitated using a language-changing framework.

Furthermore, the user interface development involved the use of HTML, JavaScript, and jQuery to create an intuitive and interactive experience for users. Deployment of the system was carried out using Docker in Google Cloud Platform (GCP), ensuring scalability and reliability, while the user interface was deployed separately using Nginx in GCP for optimal performance.

With this context in mind, let's explore the architecture of the Sinhala Language Banking Chatbot System, highlighting its components and interactions to achieve its objectives effectively.

3.4.1 User Interface Interaction

The process starts in the user interaction system architecture when a user uses a web-based user interface made with HTML, jQuery, and JavaScript to start a discussion with the chatbot. With this carefully thought-out interface, customers may easily enter their banking-related questions in Sinhala in an intuitive context. consumers may easily search for information or complete banking activities thanks to the simple design, which makes it easy for consumers to communicate with the chatbot and navigate the interface. In addition to streamlining communication, this architecture eliminates linguistic barriers, giving Sinhala-speaking users a smooth and effective way to connect with banking services via chatbots.

3.4.2 Rasa Chatbot Processing

The Rasa Chatbot, a Python implementation, receives the query. Rasa is a machine learning framework available as open source software that facilitates automated voice and text chats. In essence, Rasa understands what the user is requesting without having to formulate a response by interpreting the user's input to determine the query's intent.

3.4.3 Translation of Query to English

Once Rasa has ascertained the purpose of the user's Sinhala query, the Google Translate API is crucial in the user interaction system design as it converts the Sinhala query to English. The complex language processing models that produce the chatbot's response are written in English, therefore this translation phase is essential. This system architecture guarantees that users can converse in their preferred language while still taking advantage of the sophisticated natural language understanding and response generation capabilities of models trained primarily in English by bridging the language gap between the user's input and the chatbot's processing capabilities. Intent recognition, answer generation, and translation services are all seamlessly integrated to provide a cohesive system that improves accessibility and user experience in multilingual chatbot conversations.

3.4.5 OpenAI API for Response Generation

After Rasa ascertains the purpose of the user's Sinhala query, the query is translated from Sinhala to English via the Google Translate API. This process is required because the sophisticated language processing models that produce the chatbot's response are English-speaking. By bridging the language gap, this translation guarantees that the user's request may be appropriately understood and handled by the system architecture's later tiers. The translated question is then sent to OpenAI's API, which uses advanced LLMs to produce an answer relevant to banking. At this juncture, OpenAI's models apply their extensive knowledge base and nuanced understanding of language to produce responses that are not only accurate but also contextually relevant, thereby ensuring the chatbot's interactions are both meaningful and helpful to the user.

3.4.6 Integration with Banking Site Data

The system's ability to combine site-specific data with banking query prompts is essential for providing solutions that are customized to the particular financial environment under consideration. This vital integration guarantees that the chatbot will be able to deliver precise information specific to a given banking service or institution. In order to address each user's unique wants and inquiries, the chatbot must be configured in this way. This will improve the efficacy and relevancy of the conversations in the financial context.

3.4.7 Translation of Response to Sinhala

The English response produced by OpenAI's API is retranslated into Sinhala using the Google Translate API in order to preserve the chatbot's accessibility and guarantee usability for users who speak Sinhala. In order to provide a seamless and inclusive user experience, it is imperative that consumers receive the financial information and help they require in their native tongue. The technology successfully closes the linguistic gap by doing this, bridging the gap between the user base's linguistic preferences and sophisticated AI-driven financial advise.

3.4.8 Delivery of the Response

The system makes sure that the user experience is flawless at the last stage of the Delivery of the Response architecture by putting the translated Sinhala response right on the user interface. With this critical phase, the interaction loop is successfully closed and users are given the necessary banking information in a language that they can easily access and comprehend. By using this strategy, the design ensures that users receive pertinent help in a timely manner, hence improving their overall experience with the banking chatbot service. It also highlights the significance of linguistic inclusivity. This is illustrated in below Figure 3-2.

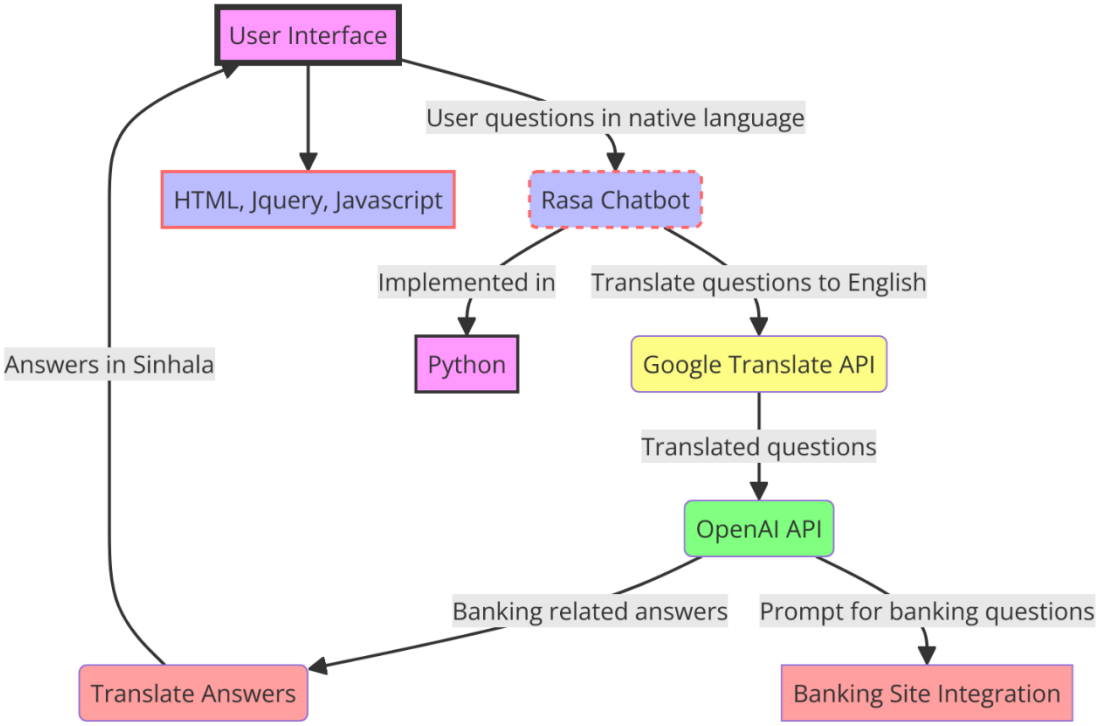


Figure 3-2 Chatbot System Architecture

3.5 User Interface Development

UI development for the Sinhala Language Banking Chatbot System involved creating a visually appealing and intuitive interface to facilitate seamless interaction between users and the chatbot. Using HTML, JavaScript, and jQuery, the UI was designed to be user-friendly, allowing users to easily input their queries in Sinhala and receive responses in the same language. The interface provided clear prompts and feedback to guide users through the conversation process, enhancing their overall experience. Additionally, considerations were made for responsiveness and accessibility, ensuring that the UI functions effectively across different devices and for users with varying levels of technological proficiency. Through thoughtful design and implementation, the UI served as a key component in enabling effective communication between users and the chatbot, ultimately enhancing accessibility to banking services for Sinhala-speaking individuals. The snapshot of UI design is denoted in below Figure 3-3.

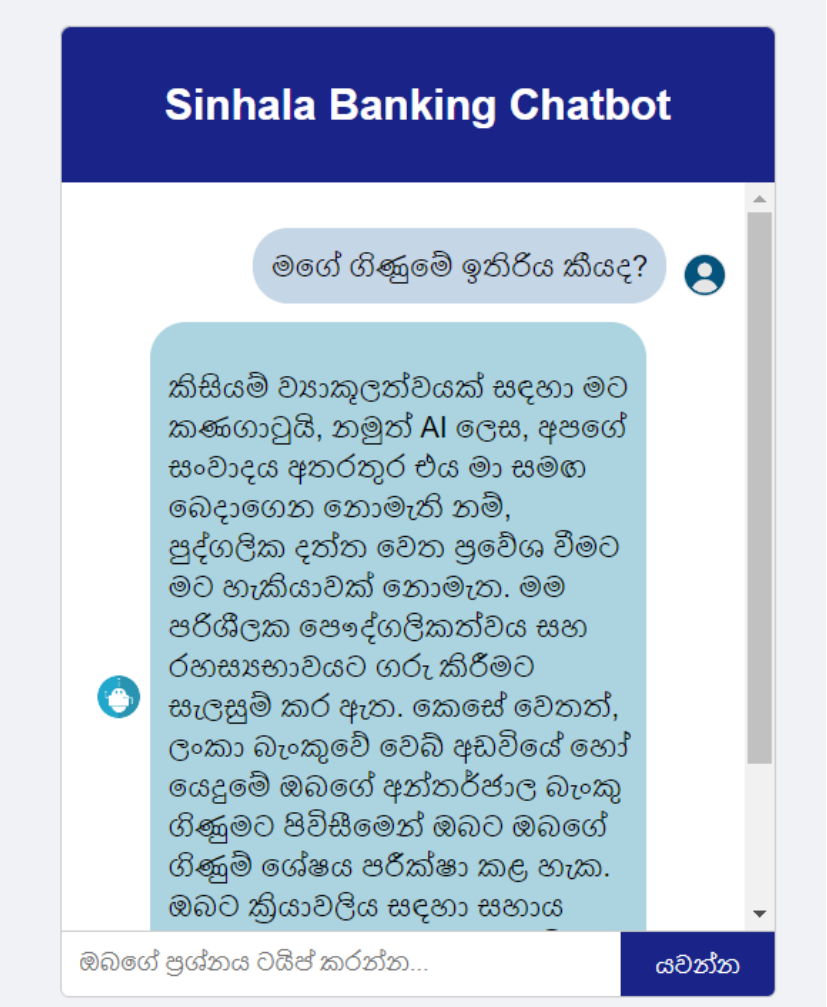


Figure 3-3 Chatbot User Interfcae

3.6 Deploy Chatbot

Deployment of the Sinhala Language Banking Chatbot System involved the process of making the system accessible and operational for users.

A number of crucial phases were engaged in the creation and implementation of the chatbot system, each selected for the unique advantages they offered to the project's success as a whole. First, servers, databases, and networking configurations were provisioned as part of the infrastructure setup process. Google Cloud Platform (GCP) was chosen as the service provider. GCP was preferred because of its stability, which ensured continuous operation without noticeable downtime, scalability, which allowed the chatbot to accommodate variable user loads, and ease of management, which reduced the administrative cost(*Containers on Compute Engine | Compute Engine Documentation | Google Cloud*, no date). The chatbot application and its related components were containerized using Docker after the infrastructure was set up. Docker made a strategic choice by streamlining the management of application dependencies and enabling consistent deployment across a variety of settings, which improved efficiency and reduced error-proneness in the development process.
In order to meet the chatbot's operational needs, Docker containers were actually deployed using the GCP infrastructure, using the platform's stability and scalability. This made sure that consumers could always use the chatbot system, even when it was busy. Finally, Nginx was set up as the web server to provide end users with the user interface (UI). The choice to utilize Nginx was supported by its effectiveness in handling HTTP requests and serving static files, both of which improved the UI's responsiveness and performance (*How can I deploy chatbot with nginx - Rasa Open Source*, 2021). This all-encompassing strategy, which combined the infrastructure capabilities of GCP with the benefits of containerization offered by Docker and the effective request handling of Nginx, offered a strong basis for implementing a dependable and high-performing chatbot system.

Deployed URL - http://34.16.161.81:8000/

CHAPTER 4
EVALUATION AND RESULTS

The Evaluation chapter of your thesis on the "Design and Development of a Sinhala Language Banking Chatbot System" is a pivotal section that aims to scrutinize the effectiveness and impact of the developed chatbot. The research hypothesis will be introduced in this chapter, along with the problem domain (the language barrier in banking chatbots that are currently in use and do not support Sinhala speakers), which will set the stage for the evaluation. The hypothesis that integrating a banking chatbot in Sinhala, built with Rasa and improved with the LangChain framework on the Google Cloud Platform, will greatly improve the efficiency and accessibility of banking services for the Sinhala-speaking community in Sri Lanka will then be examined. This introduction states the expected research outcomes and provides context for the following evaluation techniques.

4.1 Evaluation Methodology

In the evaluation methodology section, the first topic focuses on introducing the mixed-methods approach, which combines quantitative and qualitative evaluations to provide a comprehensive understanding of the chatbot's performance.

A comprehensive process that combines quantitative and qualitative evaluations, the mixed-methods approach provides a comprehensive assessment of a chatbot's efficacy. This methodology recognizes that evaluating a chatbot's functioning requires considering a wide range of factors, from subjective user experiences to objective performance indicators. Researchers can gain a thorough grasp of the chatbot's functioning by combining quantitative metrics like accuracy rates and reaction times with qualitative data like user happiness and perceived value. The utilization of a dual method enables a comprehensive review by providing a nuanced study that encompasses multiple characteristics of the chatbot's utility, usability, and overall user experience. Quantitative metrics provide an impartial evaluation of the chatbot's performance by supplying numerical data that can be statistically examined to identify the chatbot's advantages and disadvantages in relation to aspects like response accuracy and query processing speed. Conversely, qualitative insights capture user satisfaction, the chatbot's perceived utility from the user's perspective, and they also delve into the subjective aspects of user interaction with the chatbot. Through the use of a mixed-methods methodology, researchers can combine results from many data sources to produce a comprehensive assessment that combines the accuracy of quantitative performance indicators with the breadth of qualitative user feedback. his

comprehensive evaluation method allows researchers to thoroughly understand the chatbot's impact, guiding iterative improvements to enhance the system based on both the rich insights of user feedback and the concrete benchmarks provided by quantitative analysis.

4.2 Quantitative and Qualitative Assessment

The "Quantitative and Qualitative Assessment" section of the evaluation methodology involves combining both quantitative metrics and qualitative insights to comprehensively evaluate the performance and user experience of the chatbot system.

4.2.1 Combining Quantitative Metrics and Qualitative Insights

This section combines qualitative information from user comments and surveys with quantitative metrics like accuracy rates, response times, and user engagement indicators. Quantitative metrics provide numerical data that can be statistically examined and provide objective measures of the chatbot's performance. For instance, response times gauge how quickly the system generates responses, whereas accuracy rates gauge how well the chatbot comprehends user inquiries and responds accordingly.

Conversely, qualitative insights offer a subjective assessment of users' experiences, satisfaction levels, and chatbot's perceived worth. This comprises input from users obtained by questionnaires, interviews, or open-ended inquiries; these methods provide insightful information on the opinions, preferences, and problems of users. Researchers can obtain a more comprehensive picture of the chatbot's success by integrating quantitative measurements and qualitative insights, which capture both objective performance indicators and subjective user experiences.

4.2.2 Methodology for Data Collection and Interpretation

The approach for collecting and analyzing both quantitative and qualitative data is described in this section. In the case of quantitative measurements, information from the chatbot system may be automatically gathered and evaluated using statistical techniques, for example, through log files or analytics tools.

Qualitative data can be gathered in an organized or semi-structured manner, and it can be evaluated thematically or using qualitative coding techniques. Qualitative data includes user feedback via surveys and interviews. Finding patterns, trends, and correlations in data requires combining information from qualitative and quantitative sources. In order to confirm findings and

reach solid judgments regarding the functionality and user experience of the chatbot, this may include triangulating data from several sources.

4.2.3 Incorporating User Feedback

The evaluation method greatly benefits from user feedback, which offers insightful information about users' preferences, views, and interactions with the chatbot. Through surveys, interviews, or direct encounters with the chatbot, users can provide ratings, comments, ideas, and grievances. Through the integration of user feedback into the evaluation procedure, researchers are able to discern the chatbot's advantages and disadvantages, rank the areas in need of development, and make well-informed choices to optimize the system's functionality and user experience.

4.3 Experimentation and User Experience

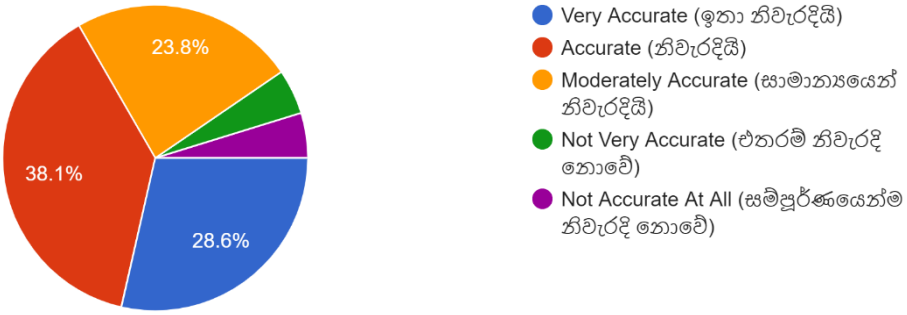
In assessing the performance and user experience of the Sinhala Language Banking Chatbot System, user feedback played a pivotal role. As part of the evaluation process, user feedback was collected to assess the effectiveness and user experience of the chatbot. A Google Doc form was created to gather direct feedback from users regarding their interactions with the chatbot. This form included structured questions meticulously crafted to evaluate various aspects of user satisfaction, usability, and perceived value. Through the utilization of this structured approach, users were prompted to provide insights into their interactions with the chatbot, enabling a comprehensive assessment of its performance. This feedback mechanism served as a valuable asset in understanding users' perceptions, preferences, and challenges encountered during their engagements with the chatbot. As a result, this integrated approach provided holistic insights into both the technical performance and user-centric aspects of the chatbot system.

Google Doc Form:
https://docs.google.com/forms/d/e/1FAIpQLSd0LqMZK3zk86K_uiAR6PlnotGYTN96fEQCtRJqdWvZpKLXZA/viewform

User Feedback :
https://docs.google.com/spreadsheets/d/15VXdLWp0lF5t2tib1UR5a_OjW80insVAk0WKvFw_qxE/edit?usp=sharing

1. How accurate were the chatbot's responses to your banking queries? (ඔබේ බැංකු විමසීම් සඳහා චැට්බොට්ගේ පිළිතුරු කොපමණ නිවැරදි විය?)

21 responses

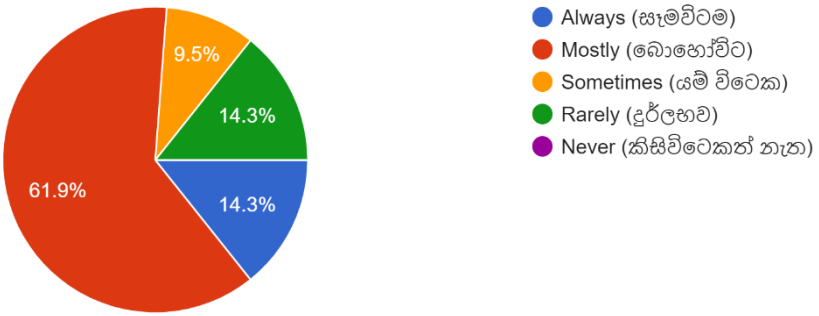


2.How effective was the chatbot in resolving your queries? (ඔබේ ප්‍රශ්න විසඳීමට චැට්බොට් කොපමණ කාර්යක්ෂම විය?)

21 responses

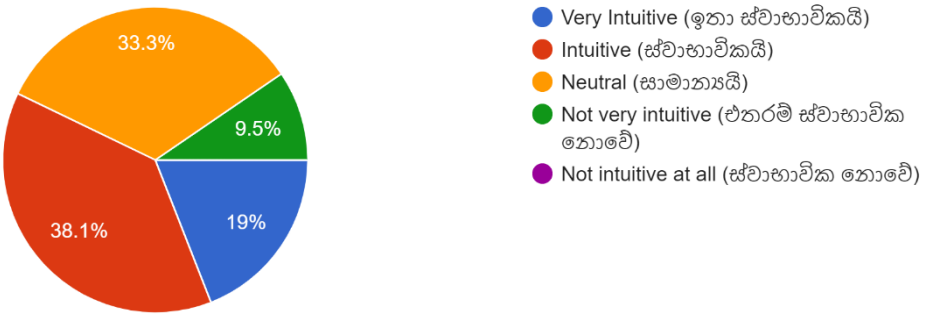
3. Was the chatbot able to understand your questions accurately? (ඔබේ ප්‍රශ්න චැට්බොට් හරියටම තේරුම් ගත්තාද?)

21 responses



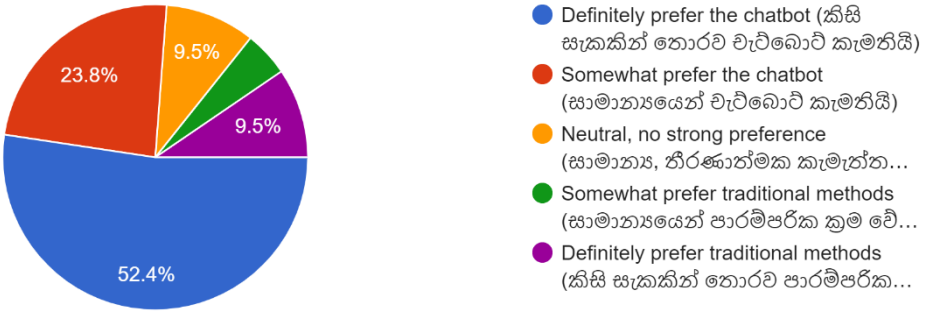
4.How intuitive did you find the chatbot's user interface? (චැට්බොට්ගේ පරිශීලක අතුරුමුහුණත කොපමණ ස්වාභාවික බවක් දක්වනවාද?)

21 responses



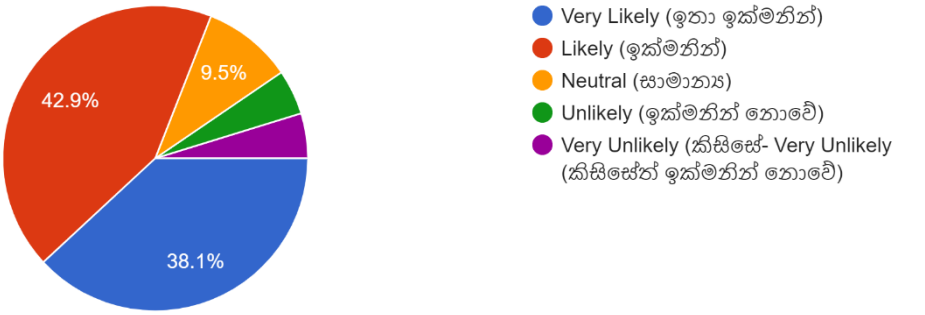
5.Would you prefer using this chatbot over traditional banking methods? (පාරම්පරික බැංකු ක්‍රම වේදයන්ට වඩා මෙම චැට්බොට් භාවිතය ඔබ කැමතිද?)

21 responses



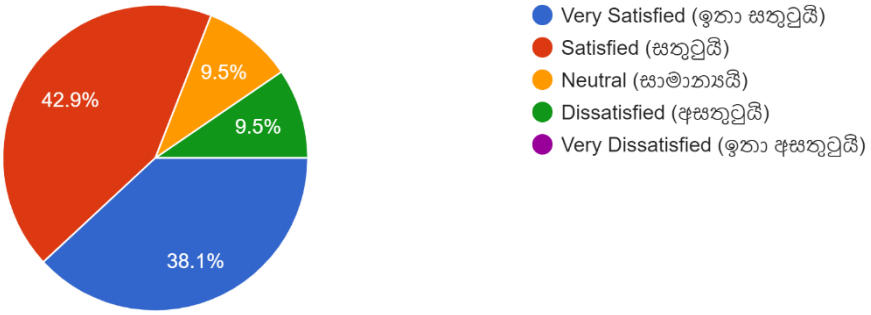
6. How likely are you to recommend our chatbot to others? (අපගේ චැට්බොට් වෙත අන් අයට ඔබ කොපමණ ඉක්මනින් නිර්දේශ කරන්නේද?)

21 responses



7.How would you rate your overall experience with the chatbot? (සම්පූර්ණ අත්දැකීම ඔබ කෙසේ තක්සේරු කරන්නේද?)

21 responses



8. Additional Comments or Suggestions

7 responses

Chatbot very use full and give better banking experience.
කාඩ්පත අහිමි වුනා, මොකද කරන්නේ?
Good
This chatbot is very useful. The way how it quickly provides relevant information and assists with the relevant inquiries is great.
Quesries with Basic problems can be solved by Chatbot. But more complex individual problems need customer support service. Rather than waiting a lot of tike with chat bot it is much appreciated if the customer support is readily available.
Good tool for modern technology.
Great initiative for the Sinhala e-banking system.

4.4 Data Analysis and Findings

The evaluation's "Data Analysis & Findings" section compiles user feedback to evaluate the chatbot's effectiveness and user happiness. Users generally perceive the chatbot's responses to be precise and helpful in answering their banking questions, according to data analysis. Users find the chatbot to have a good comprehension of their inquiries, and they find the interface to be user-friendly. Consumers show that they prefer this cutting-edge digital solution to conventional banking techniques, suggesting that it has received favorable feedback and has the potential to be

widely adopted. Strong evidence that the chatbot has improved the banking experience for Sinhala speakers comes from their readiness to refer it to others and their general level of happiness.

4.4.1 Accuracy of Responses

Users' opinions of the chatbot's accuracy in answering banking-related queries are displayed in the first pie chart, which shows that the language model and translation processes are working well. The fact that most users gave the chatbot ratings of "Very Accurate" or "Accurate" indicates that the chatbot does a great job of using context to offer relevant information and shows how well its underlying response-generation algorithms work. By utilizing its extensive language processing capabilities, the chatbot is able to handle complex banking queries with ease, as demonstrated by the analysis of this data, which provides a comprehensive knowledge of how well it uses linguistic and contextual signals to meet customers' informational demands.

4.4.2 Effectiveness in Resolving Queries

The efficiency of the chatbot in answering user inquiries is demonstrated by the second pie chart, wherein a considerable proportion of users rank it as "Very Effective" or "Effective" when it comes to resolving banking-related queries. This overwhelmingly favorable feedback highlights the chatbot's value in giving users pertinent answers, demonstrating its effectiveness in raising user satisfaction and possibly lightening the load on conventional customer support channels in the banking industry. This measure indicates the chatbot's potential to enhance the overall customer care experience by providing a dependable and easily accessible substitute for traditional customer assistance techniques, in addition to demonstrating the chatbot's capacity to process and reply to user inquiries quickly.

4.4.3. Understanding of User Questions

The third figure illustrates how well the chatbot's NLP module performs by offering insights into how well it understands user inquiries. The majority of replies confirming the chatbot's good comprehension demonstrate how strong its NLP skills are, particularly in intent recognition and entity extraction. This encouraging comment highlights the chatbot's technological sophistication, which allows it to correctly understand the subtleties of customer requests. This functionality not only improves the user experience by guaranteeing accurate and appropriate responses, but it also showcases the sophisticated status of the chatbot's underlying artificial intelligence algorithms, which are essential for enabling a seamless and natural connection between the user and the chatbot.

4.4.5 User Interface Intuitiveness

The evaluation of the chatbot's user interface intuitiveness, depicted in the fourth chart, reveals affirmative reactions, indicating a well-designed interface that enhances usability a vital aspect of user experience. The positive feedback underscores the interface's effectiveness in simplifying user interactions, a key determinant in fostering engagement. Notably, user interface design significantly influences technology adoption, as evidenced by user feedback indicating favorable reception. This affirmation underscores the interface's pivotal role in facilitating successful user engagement and underscores the importance of prioritizing intuitive design principles in chatbot development.

4.4.6. Preference Over Traditional Banking Methods

The fifth chart gauges user preference for the chatbot over traditional banking methods. The strong preference for the chatbot indicates a shift towards digital solutions among users and highlights the chatbot's potential to modernize customer interactions in banking.

4.4.7 Likelihood of Recommending the Chatbot

The sixth chart illustrates how users' propensity to refer others to the chatbot attests to both its perceived worth and user pleasure. This is a crucial sign of the chatbot's efficacy and user acceptability.

4.4.8 Overall User Experience

Users' overall happiness with their chatbot-using experience is depicted in the final chart. The favorable bias in satisfaction levels highlights the efficacy of the chatbot and the accomplishment of incorporating conversational AI into the banking industry for the Sinhala-speaking populace.

In conclusion, the analysis of the feedback indicates a strong user affinity towards the chatbot, satisfaction with its performance, and a readiness to embrace this new technology. The findings suggest that the chatbot is well-positioned to meet the needs of Sinhala-speaking users, thereby fulfilling a significant gap in the digital banking sector. These insights should be used to iterate on the chatbot's design, focusing on continuous improvement of user experience and expanding the chatbot's capabilities.

CHAPTER 5
CONCLUSION AND FUTURE WORK

5.1 Conclusion

The culmination of this research project marks a significant stride in the arena of digital banking services for Sinhala-speaking individuals. By leveraging the Rasa framework on the GCP, combined with the dynamic capabilities of LangChain AI, this project has successfully architected a robust and versatile framework for a Sinhala Banking Chatbot. Rasa's advanced NLP features provide a solid foundation for understanding and managing dialogues. In parallel, GCP ensures the system's scalability and secure hosting, catering to a potentially vast user base with varying demand patterns. The integration of LangChain AI imparts the chatbot with the ability to generate dynamic and contextually nuanced responses, an attribute that static databases alone could not achieve.

This innovative combination has yielded a highly effective and user-friendly chatbot that competently handles a wide spectrum of banking queries in Sinhala, showcasing efficiency and precision. However, it has also illuminated the limitations and challenges inherent in translation-based response generation, particularly the errors that can arise from the translation process.

Moving forward, the development of a dedicated Sinhala LLM would represent a quantum leap in this domain. A Sinhala LLM would streamline the chatbot's operational framework, enhance the accuracy of responses by removing the translation layer, and significantly reduce the potential for errors. The direct integration of such a model would align with the global movement towards more personalized and linguistically diverse AI solutions, setting a new benchmark for chatbots in the banking sector.

5.2 Future Work

Development of a Sinhala LLM: The most significant future endeavor is the creation of a dedicated LLM for Sinhala, which would be directly integrated into the chatbot. This would enhance the system's efficiency, reduce latency, and improve the accuracy of responses by removing the translation process.

Direct Integration with Banking APIs: To further improve the chatbot's functionality, a direct integration with banking APIs could be implemented, allowing the chatbot to fetch real-time data and perform transactions, thus providing a more comprehensive service.

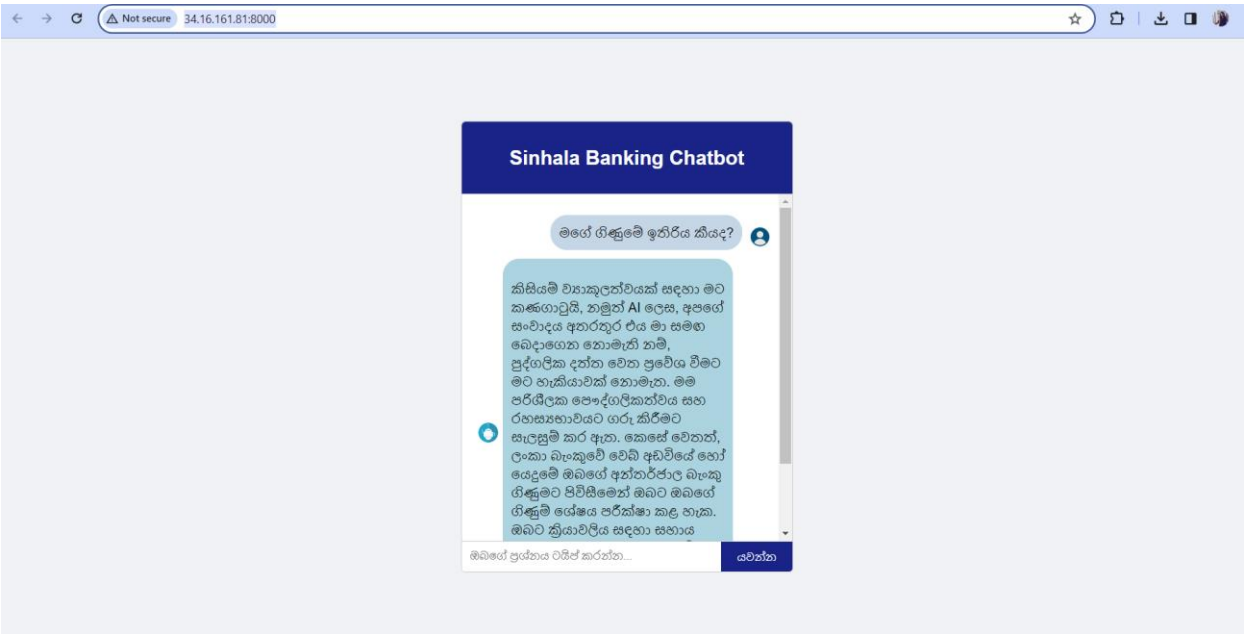
Expansion of Domain-Specific Data: Accumulating a broader set of domain-specific data would refine the chatbot's understanding and generation of responses, making it more robust against a wide variety of customer inquiries.

User Experience Enhancements: Continued improvements in the chatbot's user interface design could be explored to make it more accessible and intuitive for various user demographics.

Scalability and Deployment: Exploring scalability options to ensure that the chatbot can handle a growing number of users and queries without a decrease in performance.

APPENDICES

APPENDIX A - Deployed chatbot



APPENDIX B – nlu.yml

In order for conversational AI systems to be developed, the Rasa NLU (Natural Language Understanding) file is essential since it allows the chatbot to efficiently understand and interpret user messages. This file defines a number of languages processing tasks, including entity extraction and purpose categorization, which enable the chatbot to comprehend user inquiries and respond appropriately. The Rasa NLU file enables the chatbot to comprehend user input accurately through careful configuration and training, resulting in smooth and intuitive interactions.

version: "3.1"

nlu:

- intent: greet

examples: |

- හෙලෝ

- හයි

- intent: inquire_balance

examples: |

- මගේ ගිණුමේ ඉතිරිය කීයද?

- ගිණුමේ බලන්ස් එක කොහොමද?

- ගිණුමේ ඉතිරිය බලන්න කමතියි

- ගිණුමේ ශේෂය මොනවද?

- intent: transfer_money

examples: |

- මට සල්ලි මාරු කරන්න ඕනේ

- අනෙක් ගිණුමට සල්ලි යවන විදිය

- මට මුදල් මාරු කරන්න උදව් කරන්න

- ගිණුම් අතර මුදල් ගෙන යමට උපදෙස් දෙන්න

- intent: report_lost_card

examples: |

- මගේ බැංකු කාඩ්පත නැති වෙලා

- කාඩ්පත අහිමි වුනාමොකද කරන්නේ?

- මගේ ක්‍රෙඩිට් කාඩ්පත අතුරුදහන් වෙලා

- මගේ ඩෙබිට් කාඩ්පත හොයාගත නොහැකියි, ඉක්මනින් හඳුනුම්මට උදව් කරන්න

- intent: loan_inquiry

examples: |

- ණය ගැනීමේ ප්‍රක්‍රියාව කොහොමද?

- ණය පිළිබඳ විස්තර

- ණය ලබාගැනීමට කොන්දේසි මොනවාද?

- මට නිවඩ ඡයක් ගන්න ඕනේ, කොඉහොමද ඒ

APPENDIX C – action.py

```
from rasa_sdk import Action
from rasa_sdk.events import SlotSet
import requests
import json

# Google Cloud Translation and OpenAI API settings
google_api_key = "AIzaSyBhu4U62_O4q7168GfF4oqfEDmDduJpzh0"
openai_api_key = "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"

def translate_text(text, source_language, target_language):
    url = "https://translation.googleapis.com/language/translate/v2"
    data = {
        'q': text,
        'source': source_language,
        'target': target_language,
        'format': 'text',
        'key': google_api_key
    }
    response = requests.post(url, data=data)
    result = response.json()
    return result['data']['translations'][0]['translatedText']

def get_response_from_openai(text, session_id="my_session"):
    headers = {
```

```
        'Content-Type': 'application/json',
        'Authorization': f'Bearer {openai_api_key}'
    }

    prompt_text = (
        "The following is a customer service conversation. The assistant is helpful, "
        "creative, clever, and very knowledgeable about the Bank of Ceylon's services, "
        "products, and website information found on https://www.boc.lk/.\n\n"
        f"Customer: {text}\nAssistant:"
    )

    data = {
        'model': "gpt-4",
        'messages': [
            {'role': 'system', 'content': prompt_text},
            {'role': 'user', 'content': text}
        ],
        'temperature': 0.7,
        'max_tokens': 150
    }

    response = requests.post('https://api.openai.com/v1/chat/completions', headers=headers, json=data)

    if response.status_code == 200:
        result = response.json()
        # Get the last message from the choices array
        if 'choices' in result and len(result['choices']) > 0:
            last_choice = result['choices'][-1]  # Get the last choice from the array
```

```
            if 'message' in last_choice:
                return last_choice['message']['content'].strip()
            else:
                print("Last choice does not contain 'message':", last_choice)
                return "I'm sorry, but I couldn't fetch a response from my AI model."
        else:
            print("OpenAI response is missing 'choices':", result)
            return "I'm sorry, but I couldn't fetch a response from my AI model."
    else:
        print("Failed to fetch response from OpenAI:", response.text)
        return "I encountered an error while trying to fetch a response from my AI model."

class ActionTranslateAndRespond(Action):
    def name(self):
        return "action_translate_and_respond"

    async def run(self, dispatcher, tracker, domain):
        # Get message from the user
        user_message = tracker.latest_message['text']

        # Translate message to English
        translated_text = translate_text(user_message, "si", "en")

        # Get response from OpenAI
        response_text = get_response_from_openai(translated_text)

        # Translate response back to Sinhala
        response_in_sinhala = translate_text(response_text, "en", "si")
```

```
        # Send the response back to the user as a single string
        dispatcher.utter_message(text=response_in_sinhala)

        return []
```

APPENDIX D – docker-compose.yml

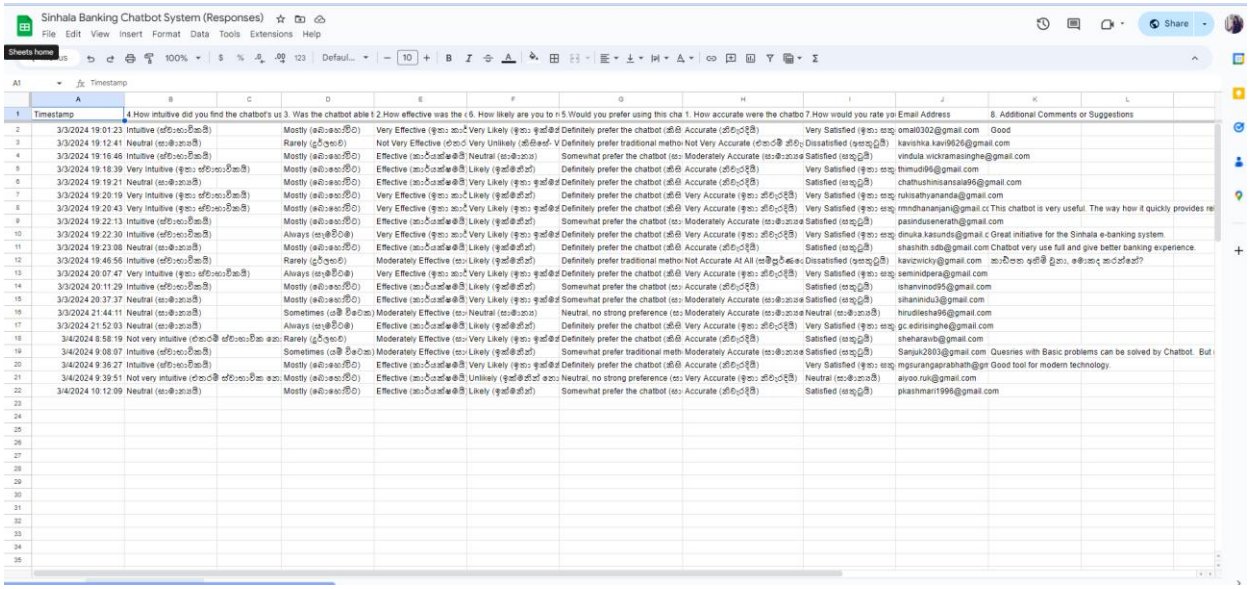
```
version: '3'
services:
  rasa:
    build:
      context: .
      dockerfile: Dockerfile
    volumes:
      - .:/app
      - ./models:/app/models
    ports:
      - "5005:5005"

  action-server:
    build:
      context: ./actions
      dockerfile: Dockerfile
    volumes:
      - ./actions:/app/actions
    ports:
      - "5055:5055"
```

```
  web:
    image: nginx:alpine
    ports:
      - "8000:80"
    volumes:
      - ./index.html:/usr/share/nginx/html/index.html
      - ./style.css:/usr/share/nginx/html/style.css
      - ./script.js:/usr/share/nginx/html/script.js
      - ./icons:/usr/share/nginx/html/icons
      - ./nginx.conf:/etc/nginx/conf.d/default.conf
    depends_on:
      - rasa

volumes:
  rasa-db:
```

APPENDIX E - User Feedbacks



REFERENCES

Amama, C. and Okengwu, U. (2023) *Smart Chatbot System for Banking using Natural Language Processing Tools*. Available at: https://doi.org/10.21203/rs.3.rs-3285543/v1.

Aslam, F. (2023) 'The Impact of Artificial Intelligence on Chatbot Technology: A Study on the Current Advancements and Leading Innovations', *European Journal of Technology*, 7, pp. 62–72. Available at: https://doi.org/10.47672/ejt.1561.

Belda-Medina, J. and Calvo-Ferrer, J.R. (2022) 'Using Chatbots as AI Conversational Partners in Language Learning', *Applied Sciences*, 12, pp. 1–16. Available at: https://doi.org/10.3390/app12178427.

Casanueva, I. *et al.* (2022) *NLU++: A Multi-Label, Slot-Rich, Generalisable Dataset for Natural Language Understanding in Task-Oriented Dialogue*.

Containers on Compute Engine | Compute Engine Documentation | Google Cloud (no date). Available at: https://cloud.google.com/compute/docs/containers (Accessed: 5 March 2024).

Dissanayake, I. *et al.* (2022) *Enhancing Conversational AI Model Performance and Explainability for Sinhala-English Bilingual Speakers*. Available at: https://doi.org/10.1109/ICAC57685.2022.10025153.

Dwivedi, Y.K. *et al.* (2023) 'Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy', *International Journal of Information Management*, 71, p. 102642. Available at: https://doi.org/10.1016/j.ijinfomgt.2023.102642.

Govindasamy, A. (2023) 'Development of a Large Language Model Chatbot for Enhanced Anime Discovery', in. Available at: https://doi.org/10.5281/zenodo.10076643.

Hettige, B. and Karunananda, A. (2006) *First Sinhala Chatbot in action*.

How can I deploy chatbot with nginx - Rasa Open Source (2021) *Rasa Community Forum*. Available at: https://forum.rasa.com/t/how-can-i-deploy-chatbot-with-nginx/49000 (Accessed: 5 March 2024).

Leocadio, P. (2023) 'Large Language Models – Machine Learning Down To Earth'. Available at: https://doi.org/10.6084/m9.figshare.22220689.

Maseke, B. (2024) 'The Transformative Power of Artificial Intelligence in Banking Client Service', *South Asian Journal of Social Studies and Economics*, 21, pp. 93–105. Available at: https://doi.org/10.9734/sajsse/2024/v21i3787.

McTear, M. and Ashurkina, M. (2024) *Transforming Conversational AI: Exploring the Power of Large Language Models in Interactive Conversational Agents*. Apress.

Permadi, I. (2023) 'Criteria Selection and Comparative Analysis of Popular Chatbot Frameworks (Dialog flow, Microsoft Bot Framework, IBM Watson Assistant and Rasa) For Implementation in Libraries: a Systematic Literature Review', *JPUA: Jurnal Perpustakaan Universitas*

Airlangga: Media Informasi dan Komunikasi Kepustakawanan, 13, pp. 94–103. Available at: https://doi.org/10.20473/jpua.v13i2.2023.94-103.

Raiaan, M. *et al.* (2023) *A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges*. Available at: https://doi.org/10.36227/techrxiv.24171183.v1.

Rasa: Developer Documentation Portal (no date) *Rasa*. Available at: https://rasa.com/docs/ (Accessed: 5 March 2024).

Wade, R. (2011) 'Try Google Translate to overcome language barriers', *BMJ (Clinical research ed.)*, 343, p. d7217. Available at: https://doi.org/10.1136/bmj.d7217.

Yusuf, A., Karaarslan, E. and Aydın, Ö. (2024) *A RAG-based Question Answering System Proposal for Understanding Islam: MufassirQAS LLM*. Available at: https://doi.org/10.36227/techrxiv.170723304.41988020/v1.

Zahra, A.M. (2022) 'Design of the use of chatbot as a virtual assistant in banking services in Indonesia', *IAES International Journal of Artificial Intelligence (IJ-AI)* [Preprint]. Available at: https://www.academia.edu/86987340/Design_of_the_use_of_chatbot_as_a_virtual_assistant_in_banking_services_in_Indonesia (Accessed: 5 March 2024).