UNIVERSITY OF MINES AND TECHNOLOGY

TARKWA

FACULTY OF COMPUTING AND MATHEMATICAL SCIENCES

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PROJECT REPORT ENTITLED

AI-BASED SPEECH TRAMSLATION TOOL FOR GHANAIAN LANGUAGES

BY

KENNEDY AKOGO KWEKU

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF BACHELOR OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING

PROJECT SUPERVISOR

....

DR HAMIDU ABDEL-FATAO

TARKWA GHANA

MONTH, YEAR

DECLARATION

I, Kennedy Kweku Akogo, declare that this project work is my own work. It is being submitted for the degree of Bachelor of Science in Computer Science and Engineering in the University of Mines and Technology (UMaT), Tarkwa. It has not been submitted for any degree or examination in any other University.

Signature of Student: ________________________________

__________ day of ________________ (year) __________

ABSTRACT

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

DEDICATION

I dedicate this project to my loving and supportive family, especially my fathers Mr Gilbert Kofi Nti and Mr Lionel Kabutey Korley. I also want to include my heartfelt dedication to my Mum ; this is for you as well. Your support has been invaluable, and I am truly grateful.

ACKNOWLEDGMENTS

I attribute my achievements to the Almighty God, who has guided and supported me throughout my journey. A heartfelt acknowledgment goes to Dr Hamidu Abdel-Fatao, my supervisor, whose unwavering dedication and brilliant insights have been instrumental in making this project a success. I am also deeply grateful to all the lecturers at UMaT who have not only nurtured my academic growth but have also provided personal support throughout my time on campus and during this project. Last but not least, I want to convey my sincere appreciation to my friends and family, who have been a constant source of emotional and physical support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASR	**A**utomatic **S**peech **R**ecognition
MT	**M**achine **T**ranslation
TTS	**T**ext to **S**peech
NLP	**N**atural **L**anguage **P**rocessing
NMT	**N**eural **M**achine **T**ranslation
BLEU	**B**ilingual **E**valuation **U**nderstudy

CHAPTER 1

INTRODUCTION

1.1 Statement of Problem

Ghana is a linguistically diverse country with over 80 local languages spoken by different ethnic groups(Azunre *et al.*, 2021c). While English is the official language, the majority of Ghanaians primarily communicate in their local languages. Unfortunately, Ghanaians face critical challenges in communicating effectively due to the fact that there's no translation medium that can translate from one Ghanaian language to another. As a result, there are issues like hindering inclusive education, impeding government services, and limiting economic opportunities for millions of Ghanaians who primarily communicate in their native tongues. In addition to this, it has made it impossible for businesses to reach and engage their local-market audience. Non-English speakers continue to have limited access to critical information, educational tools, and important services.

Some attempts have been made to address the translation needs of the Ghanaian languages, but they have faced various challenges. First and foremost, these translation models often translate only between the local Ghanaian languages and English, rather than directly between the local languages themselves. Additionally, the insufficient bilingual data and linguistic resources have hindered the accuracy and quality of translation Another is the difficulty with the complexity and ambiguity of natural languages resulting in inaccurate translations and limited coverage of the local languages (GhanaNLP, 2021).

Among the available translation systems for Ghanaian languages are Khaya Translator, Google Translate, and Microsoft Translator. However, these systems fall short in addressing the specific needs of intra-linguistic translation and often rely on English as an intermediary language, thereby failing to fully meet the communication requirements of Ghana's diverse linguistic landscape..

Due to the challenges faced by the current system, this AI-based translation tool tends to address these problems. The system when built will be able to convert the spoken language into text and convert the text back into speech. It will bridge the language gap, create room for more languages to be added and empower Ghanaians to fully participate in the social, economic, and educational spheres of the country.

1.2 Aim and Objective

The project aims to develop an innovative AI-driven speech-based translation tool specifically tailored for Ghanaian local languages, with the following key objectives:

1. To develop appropriate and culturally sound translation models for the Ghanaian languages targeting Twi and Ga
2. Employ real-time speech-based translation in order to offer easy interaction to remove language barriers.
3. To create user-friendly web and mobile interfaces for easy access to the system and easy navigation

1.3 Tools and Facilities To Be Used

1. Automatic Speech Recognition (ASR) technology
2. Machine Translation (MT) models
3. Text-to-Speech (TTS) synthesis systems
4. Datasets of audio recordings and parallel text in Ghanaian local languages
5. Web and mobile application development tools
6. Collaboration with linguistic experts and local communities

1.4 Methods Used

The methods to be used to achieve this project are:

1. Data Collection and Corpus Creation
2. Developing and training the ASR and MT models using advanced NLP techniques
3. TTS implementation for speech synthesis
4. User interface design and development

1.5 Scope of Work

This is a project to be developed into a culturally sensitive Ghanaian local language speech-translation system. This involves data gathering, model construction, user interface, community consultation, and ongoing improvement.

1.6 Work Organization

This project is structured into five key chapters. Chapter 1 encompasses the problem definition, project objectives, methods employed, necessary facilities, and project organization. In Chapter 2, you will find an in-depth exploration of the relevant literature related to the subject of study. Chapter 3 delves into the architectural design and implementation methods. Chapter 4 is dedicated to the testing phase and the discussion of results. Finally, Chapter 5 provides valuable recommendations, conclusions, and a list of references pertaining to the project.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

Language translation technologies are important for closing communication gaps among various linguistic communities, as researchers have long recognized (Hutchins and Somers, 2009). The use of artificial intelligence approaches to improve both the accuracy and efficiency of speech translation systems has drawn significant attention in recent years. These systems translate and transcribe spoken language in real-time employing neural networks, deep learning models, and natural language processing algorithms

2.2 History of Ghanaian Languages (Twi and Ga)

Ghanaian languages, notably Twi and Ga, have profound roots in Ghana's cultural legacy and fabric of society. The Twi language, also known as Akan, is a Niger-Congo language spoken by the Akan people in Ghana and the Ivory Coast. It is the most widely spoken indigenous language in Ghana, with over 6 million native speakers (Lewis *et al.*, 2015). The Twi language is part of the Kwa language family and is closely related to the Fante and Akuapem languages (Dolphyne, 1988). Historically, the Akan people migrated from the ancient Ghana Empire, and their language evolved from the Proto-Tano language (Osam, 2003). The Twi language has several dialects, including Asante Twi, Akuapem Twi, and Fante, which are mutually intelligible to varying degrees (Dolphyne, 1988).

The Ga language, on the other hand, is a Kwa language spoken by the Ga people, who are primarily located in the Greater Accra Region of Ghana. It has around 600,000 native speakers (Lewis *et al.*, 2015)). The Ga language is believed to have originated from the Guan languages, which are part of the Kwa language family (Osam *et al.*, 2011). The Ga people have a long history in the region, and their language has been influenced by the Akan languages, as well as English and other languages spoken in the area (Kropp Dakubu, 2011). The Ga language is tonal, with two distinct tones: high and low (Osam *et al.*, 2011).

Both the Twi and Ga languages played significant roles in the cultural and historical development of Ghana. The Akan people, who speak Twi, were among the earliest inhabitants of the region and established powerful kingdoms, such as the Ashanti Empire (Wilks, 1993). The Ga people, on the other hand, were influential in the coastal areas and played a crucial role in the transatlantic slave trade .

2.3 Key Challenges faced by Language Translation models

Creating AI models for translating between indigenous Ghanaian languages like Twi and Ga presents different challenges. One of the most pressing concerns is a lack of linguistic resources and data for these low-resource languages (Musah, 2019). Large parallel documents, vocabulary resources, and annotated data are in short supply for training high-quality translation models (Salawu (2021)). Data shortage might cause mistakes and biases in the resulting models.

Another problem is dealing with the characteristics and differences that exist within Ghanaian languages. These languages have complicated grammatical rules, tonal systems, and rich morphological structures that AI models may struggle to effectively capture (Doku and Nzomo, 2019). Furthermore, there are various dialects and variations within languages such as Twi, which might complicate the translation process (Musah, 2019).

Problems with language might also impede the collecting and annotation of data for model development. There is a scarcity of language experts and annotation experts fluent in Ghanaian languages, making it difficult to collect high-quality annotated data (Salawu (2021)). Cultural and environmental elements also play an important role in language usage, which AI models may struggle to understand without enough cultural understanding (Doku and Nzomo, 2019).

Furthermore, a lack of linguistic expertise and resources can delay the development and testing of AI translation models for Ghanaian languages. It can be difficult to evaluate the models' performance and assure their cultural compatibility without the assistance of local speakers and language experts (Musah, 2019). Language barriers can also impede user experience and adoption, as users may struggle with interfaces and outputs that are not tailored to their linguistic and cultural contexts (Salawu (2021)).

2.4 Significance of an AI-Based Translation System for Ghanaian Languages

There are several advantages that can be realized from the adoption of AI-based translation systems for Ghanaian languages.

Developing AI models for translating between Ghana's native languages is critical for promoting diversity in languages, preservation of culture, and economic development. Language barriers can impede access to information, education, and services, especially for underprivileged groups (Orife *et al.*, 2020). AI-powered translation systems can help close these gaps and enable speakers of local languages to fully engage in all parts of life.

One of the most significant benefits of AI translation models is their ability to improve communication and knowledge sharing across linguistic boundaries. This can improve collaboration and understanding among diverse populations, promoting unity in society and interaction between cultures (Nekoto *et al.*, 2020). Furthermore, it can increase access to educational resources, healthcare information, and government services for persons who are fluent in their native language (Salawu (2021)).

Furthermore, AI translation models can help to preserve and promote local languages, which are frequently threatened with extinction or marginalization (Doku and Nzomo, 2019). AI can assist maintain the life and use of these languages in a variety of fields, including literature, media, and education (Orife *et al.*, 2020). Creating AI translation models may potentially have economic implications. It can improve cross-border trade and commercial potential by allowing for effective communication amongst speakers of diverse languages (Nekoto *et al.*, 2020). Furthermore, it can promote the expansion of language-related enterprises such as localization, content creation, and language technology development (Salawu (2021)).

2.5 Language Translation Approaches

Several techniques for building translation models for local languages have been investigated, each with its own set of discoveries and contributions. We can divide these approaches into the following sub-headings:

2.5.1 Statistical Machine Translation (SMT)

Initial attempts to develop translation models for local languages usually employed statistical machine translation (SMT) techniques. SMT models use parallel corpora and statistical approaches to learn translation patterns and possibilities . Doku and Nzomo (Doku and Nzomo, 2019) conducted research into the challenges of developing SMT systems for African languages, including data scarcity and language complexity. They discovered that, while SMT systems performed reasonably well for resource-rich languages, their accuracy was limited for low-resource languages due to insufficient parallel data. Their findings underlined the vital importance of significant data gathering efforts and the creation of massive parallel corpora to increase SMT performance in these scenarios. Musa (Musah, 2019) investigated the usage of SMT for translating Ghanaian languages including Twi and Ga. The highlighted the significance of extensive information gathering and preprocessing for handling these languages' unique linguistic characteristics, such as tone systems and complicated vocabulary. (Musah, 2019) proved that even with minimal data, careful preparation and modification of SMT approaches could produce useful translation results, although usually less accurate than translations into more widely spoken languages. Siminyu *et al.*, 2019 investigated the usage of SMT in Kenyan languages, highlighting comparable difficulties of data scarcity and linguistic complexity. They proposed that using community-driven data gathering and existing multilingual texts in informal sectors could help to reduce the data scarcity problem.Overall, while SMT has demonstrated potential in translating local languages, its practicality has been severely limited by a scarcity of large and high-quality independent corpora. In addition, the unique linguistic characteristics of African languages, such as tone variation and complicated morphology, present additional obstacles that require customized methodologies and preprocessing procedures

2.5.2 Neural Machine Translation (NMT)

With the rise of deep learning, Neural Machine Translation (NMT) has become a common method for creating translation models. NMT models use neural networks to directly learn the mapping between source and target languages, resulting in higher performance than standard SMT systems.

Nekoto (Nekoto *et al.*, 2020) used collaborative study to create NMT models for multiple African languages, including Yoruba and Igbo. The study they conducted stressed the need of incorporating language specialists and native speakers in the data collecting and evaluation process, which leads to better translation results. Salawa (Salawu (2021)) studied the application of transfer learning and data augmentation strategies to improve NMT performance in low-resource African languages. Their findings revealed that using high-resource languages and creating synthetic data could greatly improve translation quality in low-resource contexts. (Adebara et al,2020) investigated the usage of transformer models for NMT in African languages. They discovered that transformer-based NMT systems outperform earlier models, especially when combined with strategies such as transfer learning and multilingual training. Their results highlighted the importance of establishing innovative neural architectures, as well as the potential benefits of training models on many related languages at the same time. Overall, NMT has showed substantial potential to improve translation quality in local languages. However, its usefulness is limited by the availability of high-quality training data. Cooperation among local populations, language experts, and advanced neural approaches is critical for overcoming these problems and improving the performance of NMT systems for low-resource languages.

2.5.3 **Hybrid and Multi-Source Approaches**

Recognizing the strengths and limits of various approaches, researchers investigated hybrid and multi-source strategies for constructing translation models. Orife (Orife *et al.*, 2020) used a combination of rule-based, SMT, and NMT approaches to develop an Igbo language technology package. Their findings pointed out the significance of utilizing a range of approaches and information sources to solve the many problems involved with low-resource language translation. They proved that combining these approaches can improve the overall translation quality and durability of the system. Specifically, rule-based techniques aided in dealing with grammatical complications, SMT provided statistical reliability, and NMT provided improved contextual awareness using neural networks.

Ahmad (Ahmad *et al.*, 2022) developed a multi-source method that used numerous related languages to increase NMT performance for a low-resource target language. Their findings demonstrated that adding data from related languages can greatly improve translation quality, particularly when dealing with limited data for the target language. The model could train more effectively and produce more accurate translations by taking advantage of linguistic similarities and shared structures between related languages. This strategy proved especially useful when direct translation data for the target language was scarce, demonstrating the power of multi-source strategies in overcoming data limitations. (goyvaerts, 2021 investigated the use of hybrid approaches in translating Central African languages. They discovered that combining linguistic concepts with machine learning models allowed them handle complicated grammatical and structural aspects common to these languages. Their findings highlighted the significance of customized hybrid systems that can adapt to the unique linguistic unique features of the target languages.Overall, hybrid and multi-source methods showed great potential for improving translation quality in low-resource languages. These approaches, which blend different methodologies and use data from related languages, may reduce the challenges caused by data scarcity and linguistic difficulty, resulting in more effective and dependable translation systems.

2.5.4 **End-to-End Speech Translation**

End-to-end speech translation systems remove the requirement for separate components by combining voice recognition and translation in a single neural network architecture (Chorowski *et al.*, 2014). These technologies provide quicker and more smooth translation experiences by converting vocal input into translated output.Chorowski et al (Chorowski *et al.*, 2014) innovated the combination of voice recognition and translation into a single model. Their solution used recurrent neural networks (RNNs) to receive spoken input and output translated text, proving that a single design could handle both tasks successfully. Vaswani et al Vaswani *et al.*, 2017 provided the transformer model, a widely adopted design that significantly enhanced the capabilities of neural machine translation systems, including end-to-end speech translation. The transformer model employs self-attention methods to capture relationships between words in a sentence more effectively than RNNs, resulting in improved performance in both text and speech translation tasks. The use of transformer models for end-to-end voice translation has resulted in significant improvements in accuracy and processing speed. (Jia et al,2019) also conducted work that extended the transformer model for end-to-end voice translation. They showed that using a transformer-based architecture might result in cutting-edge performance on a variety of benchmark datasets. Their research shown that transformer models could efficiently handle the complexity of direct speech-to-text translation, including different accents and speech patterns.Furthermore,(Di Gangi et al,2019) investigated the application of sequence-to-sequence models for end-to-end voice translation.

They investigated various architectures and training methodologies, and discovered that including auxiliary tasks like automatic voice recognition (ASR) and machine translation (MT) into a multitask learning framework could increase the translation system's overall performance.

2.6 Review of Related Works

The integration of NLP and machine learning has offered innovative solutions for language barrier. This section explores the research in this domain, highlighting key findings, methodologies, and contributions.
Table 2.1 summarizes the methodologies and findings of some of such research works.

Table 2.1 Review of Related Works

Author	**Methodology**	**Findings**	**Proposed Solution**
Gyasi and Schlippe, 2023	10,708 phrase pairings were used to generate a parallel corpus of Twi-French-English that was assessed using the BLEU, Azunre-BLEU, and SacreBLEU metrics. The top French-Twi system scored 0.81, and the top Twi-French system scored 0.76. Comparing it to Google Translate, an improvement of 7% was made. Future research will add more African low-resource languages, broaden the corpus, and look at language-specific algorithms.	This paper has the successful creation of a parallel Twi-French-English corpus, the effective corpus splitting for training and evaluation, the utilization of all the evaluation metrics, the development and fine-tuning of machine translation systems, the performance of these systems, the comparison with Google Translate, and directions for future research.	Includes extending the corpus with complex sentences from various domains and investigating language-specific algorithms for pre-processing to address Twi language peculiarities. It should also provide a web interface for broader access to the machine translation systems and expand the corpus with more African low-resourced languages.

Azunre *et al.*, 2021b	The study involved gathering and curation of existing datasets and a parallel machine translation corpus from English to Akuapem Twi. Knowledge distillation and DistilBERT were used to build condensed models, and transformer-based language models were trained and improved. Cosine similarity evaluations and correlation with human evaluations were two evaluation metrics.	The project developed transformer-based language models, which afterwards were optimized for Twi context text embeddings using datasets such as BERT, ALBERT, and RoBERTa. Smaller versions have become possible without sacrificing performance by using distillBERT and knowledge distillation. Through the use of the JW300 corpus for integration and training, the correlation with human evaluations was improved.	Includes fine-tuning on cleaner and less biased Twi data, developing a Twi version of the labeled GLUE Dataset for various tasks, training ALBERT for Twi and comparing it with existing models, exploring unsupervised spell-checking and Named Entity Recognition methods based on the models, and developing a Twi version of the GPT causal transformer-based text generator.

Azunre *et al.*, 2021c	NLP Ghana uses crowdsourcing and human correction of machine-translated data, together with the creation of embeddings for Ghanaian languages, as methods for gathering data. Large training datasets for Akan, Ewe, Ga, and other Ghanaian languages have also been a focus of the group, which has enlisted volunteers with a range of abilities to help develop and carry out its mission. In order to supplement its internally generated data, NLP Ghana has also worked with Zindi Africa and made use of pre-existing multilingual datasets including the Bible and JW300. Additionally, the group has made its statistics and models publicly available to support ongoing study on Ghanaian languages.	The study outlines difficulties, such as the absence of evaluation datasets, in developing NLP applications for Ghanaian languages. NLP Ghana seeks to enhance resources for low-resource languages, such as Akan, Ga, and Ewe, by assembling a group of scholars and specialists and emphasizing financing and support from stakeholders.	Expanding data collection to include audio data for Ghanaian languages. Annotating datasets is planned to enhance NLP tasks like Named Entity Recognition and Part of Speech tagging. Significant funding is sought to acquire quality data for NLP research. Collaboration with local and international entities is key to advancing NLP applications for Ghanaian languages and beyond.

Yamoah, 2023	It examines how text-to-speech (TTS), Natural Language Processing (NLP), and Artificial Intelligence (AI) are being used in Africa, particularly for languages like Twi, Hausa, Wolof, and Nigerian Pidgin English. Methods include building voice corpora, applying neural architectures to ASR systems, and employing models like ABENA for NLP tasks in Twi. It discusses issues including code-switching and data scarcity and offers creative solutions for Ghana's visually impaired population. The paper outlines these technologies' advancements.	The study highlights the successful development of ASR systems for African languages like Hausa, Wolof, and Nigerian Pidgin English using various models. Additionally, the application of NLP models such as ABENA for the Twi language shows promising results despite encountering religious bias in the training data. Innovative approaches are proposed to address challenges in implementing ASR, NLP, and TTS technologies in Africa, particularly focusing on aiding visually impaired individuals through a combination of computer vision, NLP, and TTS technologies. .	Includes leveraging the Wolof speech corpus for NLP and TTS tasks to enhance language processing capabilities. Addressing challenges like data scarcity, code-switching, and biases in training datasets is crucial for advancing the implementation of ASR, NLP, and TTS technologies in Africa.

Azunre *et al.*, 2021a	Using 50,000 English sentences from tatoeba.org and crowdsourced Twi translations using a Google Form poll, the English-Twi Parallel Corpus for Machine Translation was established. It was the responsibility of ten researchers to confirm and edit the translations. First translations were produced by the OPUS-MT pre-trained model and then refined by the Adam optimizer. The BLEU score was utilized to assess the translation quality, indicating the careful procedure that was followed in creating and enhancing the parallel corpus for machine translation.	The results of using a transformer-based translator to produce preliminary Akuapem Twi translations that were then improved by native speakers are presented in this study. To improve accuracy, human intervention was essential in the grading, verification, and correction of translations, building a sizable training corpus with the goal of ongoing translation improvement and cultural sensitivity. After fine-tuning, the obtained BLEU score of 0.720 showed a 3.75% improvement over the initial model, offering accurate translations for ideas that are universal.	In order to ensure correctness and get rid of translationese, human intervention in scoring, confirming, and revising translations should be highly valued. Machine translation models be continuously modified using the dataset in order to improve performance and yield accurate translations. It is recommended to reduce noise in the translations by separating the Asante and Akuapem dialects inside the data.

BIBLIOGRAPHY

Ahmad, W., Bai, H., Yu, J., Pino, J., and Black, A. (2022). "Combining Related Languages for Low-Resource Neural Machine Translation". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pp. 3750–3765. URL: https://aclanthology.org/2022.acl-long.258.pdf.

Azunre, P., Osei, S., Addo, S., Adu-Gyamfi, L. A., Moore, S., Adabankah, B., Opoku, B., Asare-Nyarko, C., Nyarko, S., Amoaba, C., Appiah, E. D., Akwerh, F., Lawson, R. N. L., Budu, J., Debrah, E., Boateng, N., Ofori, W., Buabeng-Munkoh, E., Adjei, F., Ampomah, I. K. E., Otoo, J., Borkor, R., Mensah, S. B., Mensah, L., Marcel, M. A., Amponsah, A. A., and Hayfron-Acquah, J. B. (2021a). *English-Twi Parallel Corpus for Machine Translation*. arXiv: 2103.15625 [cs.CL].

Azunre, P., Osei, S., Addo, S., Adu-Gyamfi, L. A., Moore, S., Adabankah, B., Opoku, B., Asare-Nyarko, C., Nyarko, S., Amoaba, C., *et al.* (2021b). "Contextual text embeddings for twi". In: *arXiv preprint arXiv:2103.15963*.

— (2021c). "Nlp for ghanaian languages". In: *arXiv preprint arXiv:2103.15475*.

Chorowski, J., Bahdanau, D., Cho, K., and Bengio, Y. (2014). "End-to-end continuous speech recognition using attention-based recurrent nn: First results". In: *arXiv preprint arXiv:1412.1602*.

Doku, R. and Nzomo, J. (2019). "Challenges of Building Intelligent Machine Translation Systems for African Languages". In: *Proceedings of the Future Technologies Conference (FTC) 2019*, pp. 1115–1124. DOI: 10.1007/978-3-030-32962-4_86. URL: https://link.springer.com/chapter/10.1007/978-3-030-32962-4_86.

Dolphyne, F. (1988). *The Volta-Comoe Languages*. Ed. by L. Gerhardt. St. Augustin: Wissenschaftsverlag.

GhanaNLP (2021). *Khaya Translator Web App*. https://ghananlp.org/project/translatorwebapp. Accessed: January 17, 2024.

Gyasi, F. and Schlippe, T. (2023). "Twi Machine Translation". In: *Big Data and Cognitive Computing* 7.2, p. 114.

Kropp Dakubu, M. (2011). “The Central Togo Remnant Languages”. In: *The Sociolinguistics of Ghana. The Frankfurt Contribution*. Ed. by K. Lotsu. Amsterdam: John Benjamins Publishing Company, pp. 141–194.

Lewis, M., Simons, G., and Fennig, C. (2015). *Ethnologue: Languages of the World*. 18th ed. Accessed: 4 May 2024. Dallas, Texas: SIL International. URL: https://www.ethnologue.com.

Musah, A. (2019). “The Challenges of Building Machine Translation Systems for Ghanaian Languages”. In: *Journal of Computer Science and Applications* 7.1, pp. 1–8. URL: https://www.researchgate.net/publication/333801693.

Nekoto, W., Marivate, V., Matsila, T., Fasubaa, T., Kolawole, T., Fagbohunge, T., Adesanya, S., Alo, P., Adeyemi, S., Mokgesi-Mokgese, K., Sharifaith, S., Tada, S., Karani, S., and Bathseba, M. (2020). “Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages”. In: *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 2144–2160. URL: https://aclanthology.org/2020.findings-emnlp.195.pdf.

Orife, I., Kreutzer, J., Sibanda, T., Whitenack, D., Simiyu, M., Martinus, L., Akera, J., Kinuthia, J., and Abade, S. (2020). “Leveraging Language Technology for Under-Resourced Languages: A Case Study of Igbo”. In: *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 6855–6863. URL: https://aclanthology.org/2020.lrec-1.847.pdf.

Osam, E. (2003). “An Introduction to the Verbal and Multi-Verbal System of Akan”. In: *Proceedings of the Workshop on Multi-Verb Constructions. Trondheim Summer School*. Accessed: 4 May 2024. URL: http://www.ling.hf.ntnu.no/tross/trondheimtravaux.pdf.

Osam, E., Duah, R., and Blay, A. (2011). “The Ga Language”. In: *The Sociolinguistics of Ghana. The Frankfurt Contribution*. Ed. by K. Lotsu. Amsterdam: John Benjamins Publishing Company, pp. 195–238.

Salawu, S. (2021). “Low-Resource Machine Translation for African Languages”. In: *Proceedings of the 2nd Workshop on Multi-disciplinary Approaches to Code-Switching*, pp. 47–56. URL: https://aclanthology.org/2021.multidisciplinary-codeswitching.7.pdf.

Siminyu, N., Mureithi, A., and Omondi, B. (2019). “Statistical Machine Translation for Kenyan Languages: Overcoming Data Scarcity and Complexity”. In: *Proceedings of the African Conference on Computational Linguistics (AfriCL)*, pp. 45–54.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). “Attention is all you need”. In: *Advances in neural information processing systems* 30.

Wilks, I. (1993). *Forests of Gold: Essays on the Akan and the Kingdom of Asante*. Athens: Ohio University Press.

Yamoah, K. (2023). “Efforts to Apply Natural Language Processing Technologies in Africa”. In: *Science Engineering Entrepreneurship Design (SEED) Journal* 2.1.

graphicx

CHAPTER 3

SYSTEM DESIGN AND ANALYSIS

3.1 Overview

This chapter focuses on the practical execution of the project's objectives and the assessment of the system's structure. It begins with a detailed explanation of the methodology used, including data gathering for library development, Automatic Speech Recognition (ASR) and Machine Translation (MT) model development using advanced Natural Language Processing (NLP) techniques, Text-to-Speech (TTS) implementation for voice translation, and user interface design and development. Finally, the chapter assesses the effectiveness and efficiency of the implemented design, considering variables such as correctness, usability, and scalability.

3.2 Requirement Analysis

This involves creating a precise requirements specification document that serves as a blueprint, ensuring the software meets the expected functionality and quality standards. Functional and non-functional requirements are defined in the requirement analysis.

3.2.1 Functional Requirements

- The system accurately transcribes spoken words in Ghanaian local languages.
- The system translates spoken and written text between Ghanaian languages with high accuracy.
- The resulting speech from translated text is clear.
- The language models are continuously updated to incorporate new vocabulary.
- The system has web and mobile interfaces for easy access to the translation tool.
- Users can input text or speak directly for translation, with options to adjust settings.
- The system supports real-time translation with minimal latency.
- The system has quality assurance mechanisms to detect and correct translation errors, ensuring reliability.

3.2.2 Non-Functional Requirements

- The system responds quickly, with minimal delay during speech recognition and translation, and produces results within a short time.
- The user interfaces are clear and accessible across different devices and screen sizes.
- The translation tool consistently delivers accurate and stable performance across various conditions.
- User confidentiality is maintained throughout the process.
- The system handles growing workloads and user demand without sacrificing performance or reliability.
- The translation tool is cross-platform, supporting desktops, laptops, tablets, and smartphones while maintaining functionality and user experience.
- The system complies with relevant laws and standards regarding data protection, accessibility, and language localization to ensure legal and industry best practices.

3.3 Methods Used

The methods used are as follows:

3.3.1 Data Collection and Corpus Creation

Data collection is essential for developing effective ASR and MT models. During this stage, we collected a wide variety of spoken and written data in Twi and Ga. Two approaches were used:

- **Crowdsourcing**: Engaged native speakers through online platforms and local communities to contribute speech recordings and written texts.
- **Partnerships**: Collaborated with linguistic organizations, educational institutions, and community groups to access existing language resources and corpora.

Once collected, the data was annotated, transcribed, and cleaned to create high-quality corpora suitable for training the translation model.

3.3.2 Developing and Training ASR and MT Models Using Advanced NLP Techniques

This involved leveraging advanced NLP techniques to develop and train ASR and MT models specifically for Ghanaian languages (Twi and Ga). The steps included:

- **Feature Extraction**: Extracted vital characteristics from speech sounds to represent audio data in a machine learning-compatible way. We used Mel-frequency cepstral coefficients (MFCCs) to capture important features of the sound, such as its frequency distribution and overall structure.
- **Model Selection**: Selected the Transformer model architecture for both ASR and MT tasks due to its exceptional performance in various NLP tasks. The self-attention mechanism of the Transformer model effectively captures global dependencies and contextual information, making it ideal for ASR and MT applications.
- **Training**: Trained the Transformer model using collected speech and text data through supervised learning. This involved feeding the model input and output pairs and adjusting parameters to minimize prediction errors. Fine-tuned pretrained models to reflect the complex structure of Ghanaian languages by updating parameters on smaller, language-specific datasets.
- **Evaluation**: Measured model performance using metrics such as Word Error Rate (WER) for ASR and BLEU score for MT. WER assesses transcription accuracy, while BLEU score evaluates translation quality. Iterative refining based on evaluation findings improved model accuracy.

3.3.3 Implementing Text-to-Speech for Speech Recognition and Translation

The process of implementing TTS involved converting text into sound outputs.

- **Text Normalization**: Normalized punctuation, formatting, and special characters in the translated text to prepare it for translation. Numbers were transformed into words, and spelling errors were corrected.
- **Model Selection**: Selected the Tacotron 2 model for TTS because it excels at generating high-quality, natural-sounding speech. Tacotron 2 converts text into a spectrogram, which is then turned into actual speech using WaveGlow. We trained Tacotron 2 using

recordings in Twi, Ga, and other Ghanaian languages, fine-tuning it to capture the unique characteristics of these languages.

3.3.4 User Interface Design and Development

The user interface (UI) played a crucial role in making the speech-based translation tool user-friendly. We built a web application with the following key features:

- **Design and Compatibility**: Designed intuitive UI components through wireframing and prototyping to ensure a seamless user experience. The application was made compatible with both web and mobile platforms, ensuring accessibility across various devices.
- **Speech Integration and Feedback**: Integrated speech input features, allowing users to interact with the tool using spoken commands. Incorporated feedback mechanisms like progress indicators and error messages to provide real-time updates during interactions. Supported localization to accommodate multilingual content and cultural preferences.

3.4 User Roles and Permissions

Here are the user roles and permissions for the AI-powered speech-based translation tool designed for Ghanaian native languages:

Regular User

- Use the translation tool to capture and translate spoken or written text between Ghanaian languages.
- Customize settings such as preferred language pairs and speech recognition preferences.
- View translation history and save favorite translations for future reference.

Administrator

- Manage user accounts, including creating, editing, and deleting accounts.
- Configure system settings such as default languages, speech recognition models, and translation algorithms.

- Perform system maintenance tasks such as software updates, database backups, and monitoring system performance.
- Access advanced features for troubleshooting technical issues and resolving user inquiries.

Linguistic Expert

- Collaborate with the development team to provide linguistic insights and cultural context for improving translation accuracy and fluency.
- Access and evaluate linguistic data to determine language patterns, dialect variances, and areas for improvement.
- Contribute to the creation and improvement of translation models by providing feedback on language coverage, vocabulary usage, and cultural differences.

3.5 System Architecture

The system architecture of the AI-driven speech-based translation tool for Ghanaian native languages consists of interconnected parts that support seamless translation processes. The key components include modules for Automatic Speech Recognition (ASR), Machine Translation (MT), and Text-to-Speech (TTS) synthesis. The ASR module converts spoken words in Ghanaian languages into text, which the MT module then translates into the target language. The TTS module then produces the translated text as speech, ensuring clarity and naturalness. The system also includes a user interface layer, providing simple web and mobile interfaces for user interaction. These interfaces allow users to enter spoken or written content for translation, adjust settings, and access additional features.

Behind the scenes, the system trains and refines the ASR and MT models using audio recordings and parallel text datasets in Ghanaian languages, employing advanced NLP approaches. Collaboration with linguistic experts and local groups ensures that translations are culturally relevant and grammatically accurate. Overall, the system architecture aims to provide accurate, dependable, and accessible translation services across multiple platforms and devices.

3.6 Overall System Design

The figures below illustrate the flow of the process.

The Following are the Use Case diagrams for the various roles.

Flowchart for the Administrator

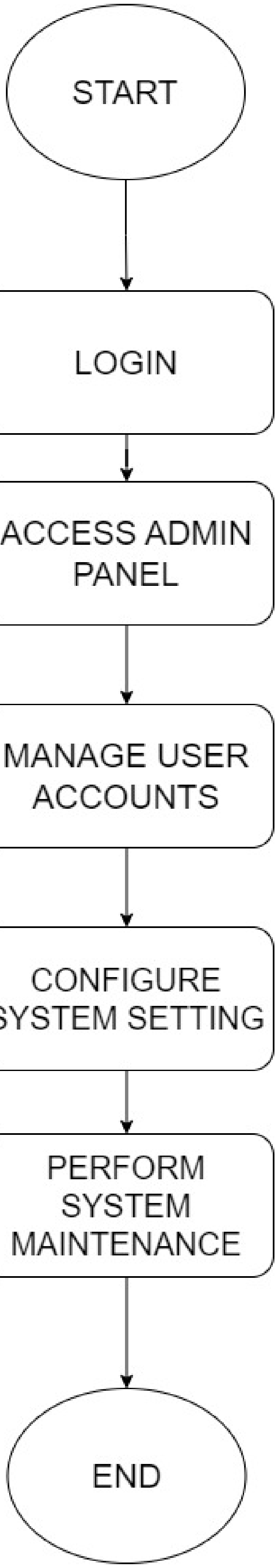

Figure 3.1 Flowchart for the Administrator

Flowchart for the Regular User

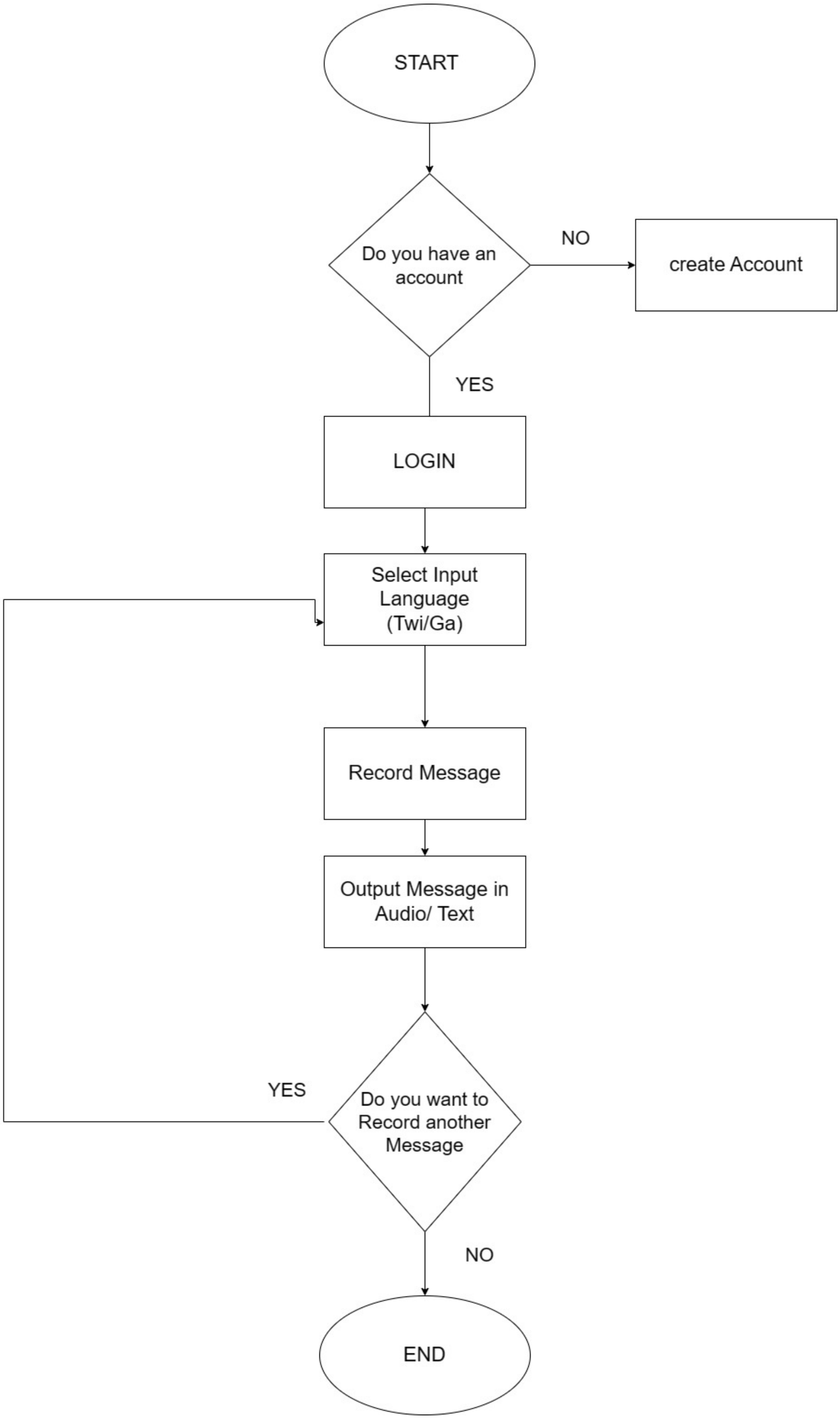

Figure 3.2 Flowchart for the Regular User

Flowchart for the Linguistic Expert

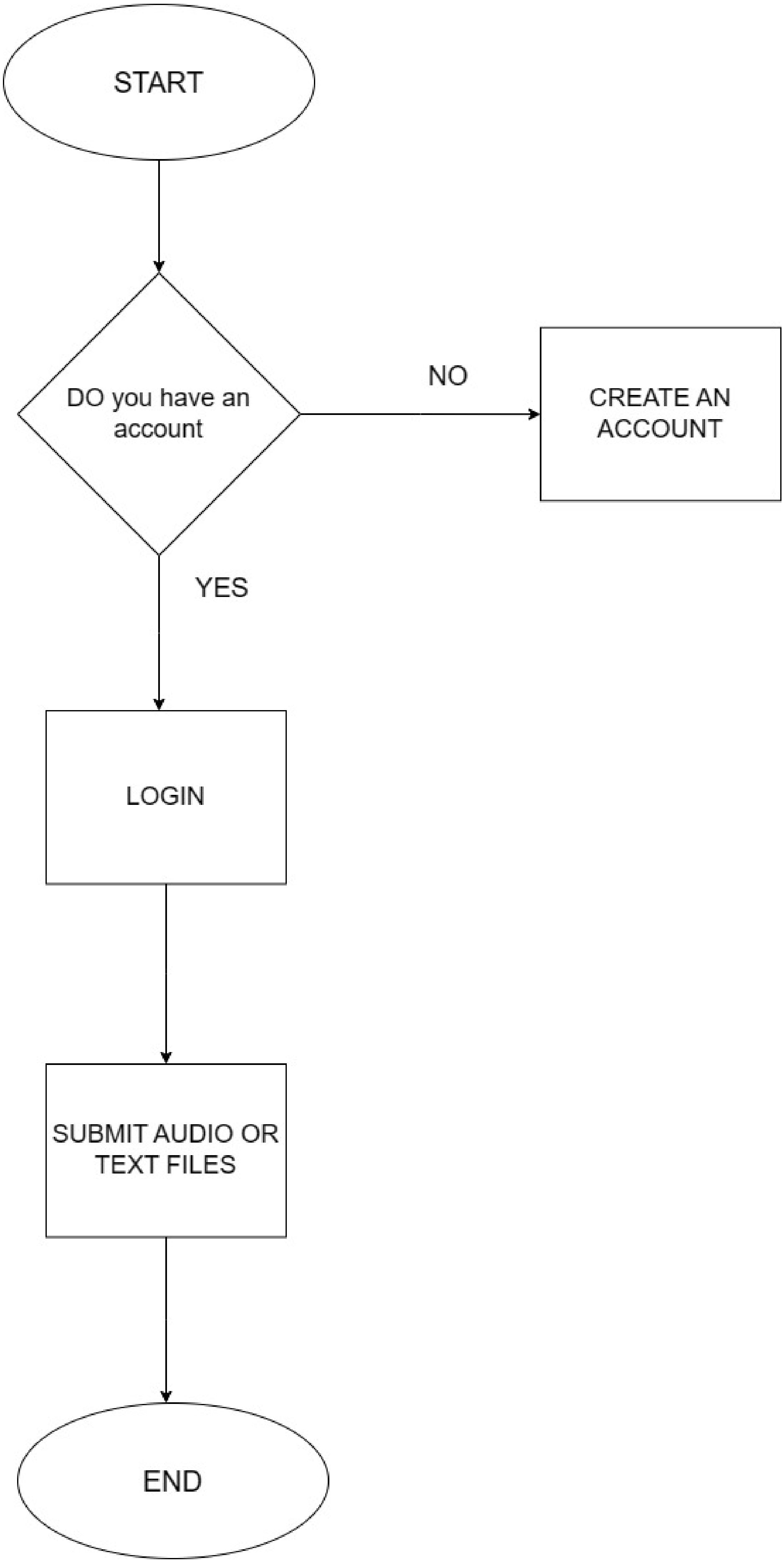

Figure 3.3 Flowchart for the Linguistic Expert

Use Case for Administrator

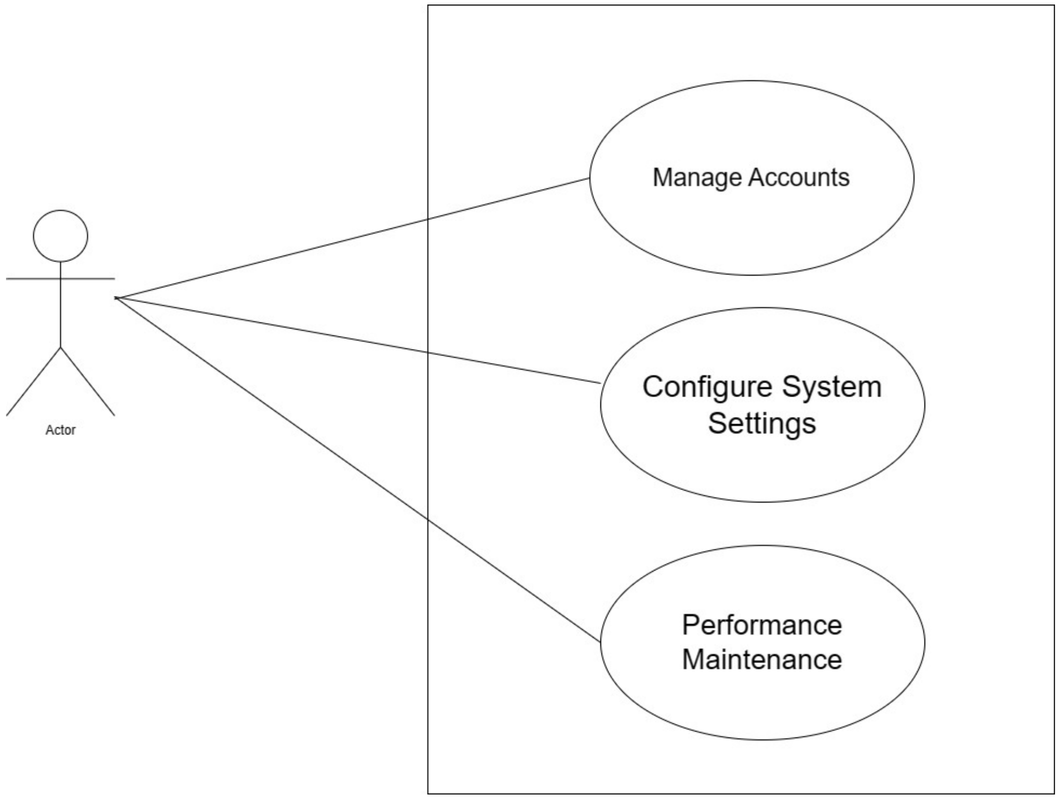

Figure 3.4 Use Case for Administrator

Use Case for Regular User

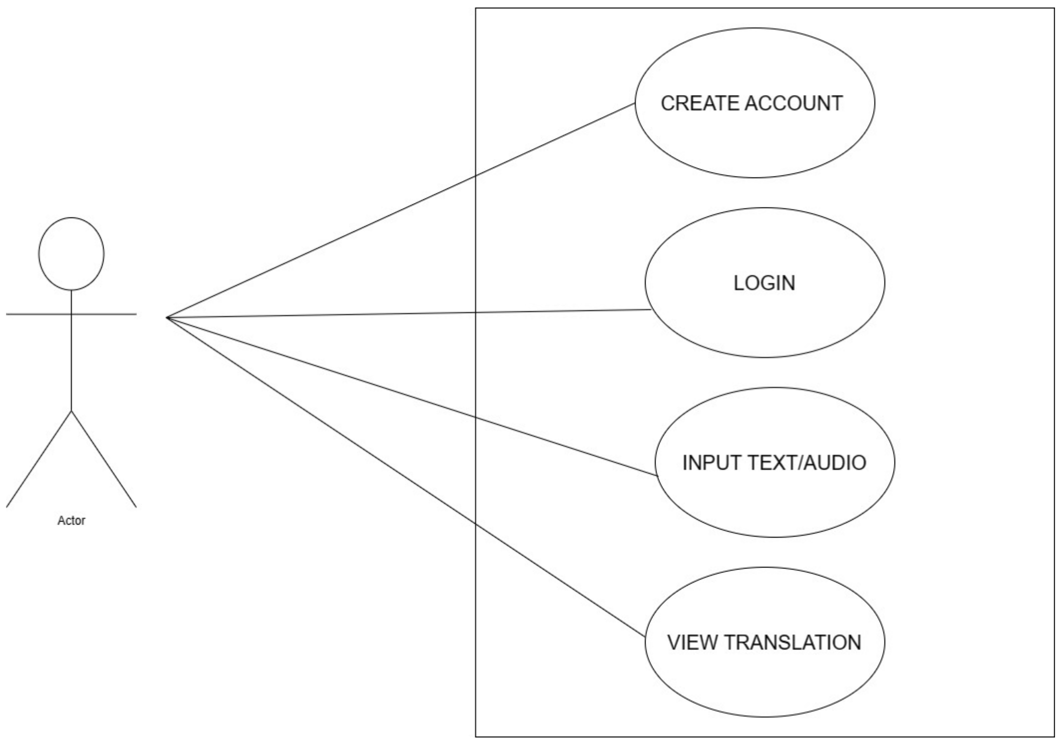

Figure 3.5 Use Case for Regular User

Use Case for Linguistic Expert

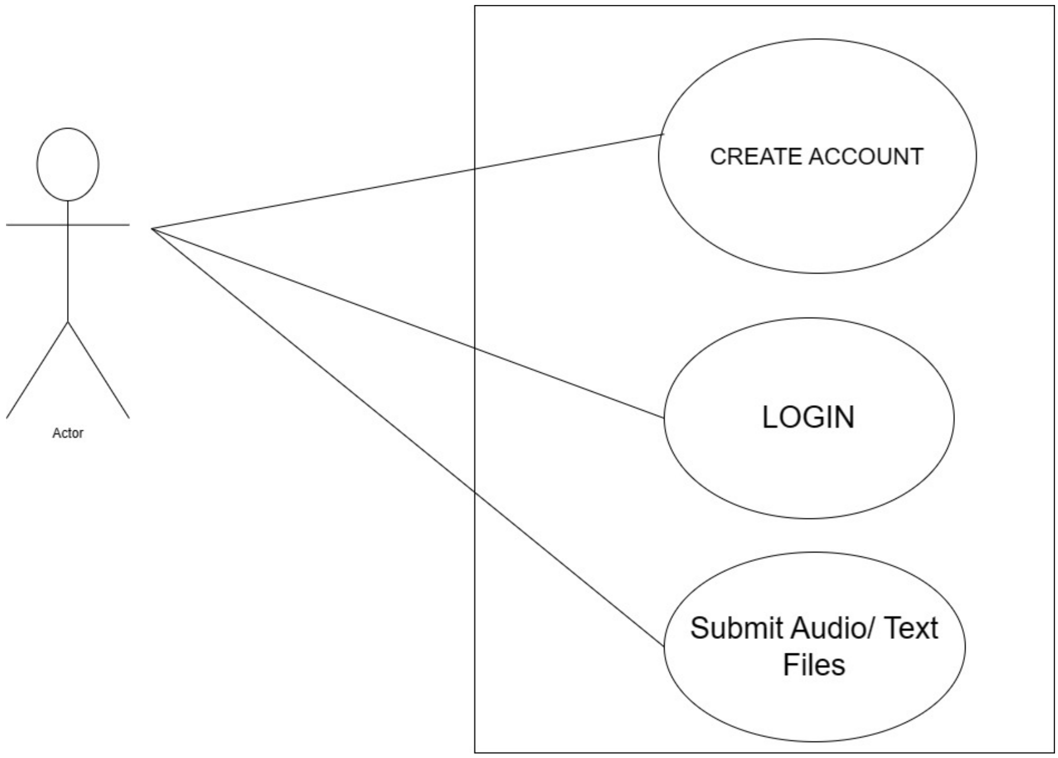

Figure 3.6 Use Case for Linguistic Expert

CHAPTER 4

CHAPTER TITLE HERE

4.1 FIRST SECTION TITLE GOES HERE

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

4.1.1 Your subsection

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

5.2 Recommendations

REFERENCES

Ahmad, W., Bai, H., Yu, J., Pino, J., and Black, A. (2022). "Combining Related Languages for Low-Resource Neural Machine Translation". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pp. 3750–3765. URL: https://aclanthology.org/2022.acl-long.258.pdf.

Azunre, P., Osei, S., Addo, S., Adu-Gyamfi, L. A., Moore, S., Adabankah, B., Opoku, B., Asare-Nyarko, C., Nyarko, S., Amoaba, C., Appiah, E. D., Akwerh, F., Lawson, R. N. L., Budu, J., Debrah, E., Boateng, N., Ofori, W., Buabeng-Munkoh, E., Adjei, F., Ampomah, I. K. E., Otoo, J., Borkor, R., Mensah, S. B., Mensah, L., Marcel, M. A., Amponsah, A. A., and Hayfron-Acquah, J. B. (2021a). *English-Twi Parallel Corpus for Machine Translation*. arXiv: 2103.15625 [cs.CL].

Azunre, P., Osei, S., Addo, S., Adu-Gyamfi, L. A., Moore, S., Adabankah, B., Opoku, B., Asare-Nyarko, C., Nyarko, S., Amoaba, C., *et al.* (2021b). "Contextual text embeddings for twi". In: *arXiv preprint arXiv:2103.15963*.

— (2021c). "Nlp for ghanaian languages". In: *arXiv preprint arXiv:2103.15475*.

Chorowski, J., Bahdanau, D., Cho, K., and Bengio, Y. (2014). "End-to-end continuous speech recognition using attention-based recurrent nn: First results". In: *arXiv preprint arXiv:1412.1602*.

Doku, R. and Nzomo, J. (2019). "Challenges of Building Intelligent Machine Translation Systems for African Languages". In: *Proceedings of the Future Technologies Conference (FTC) 2019*, pp. 1115–1124. DOI: 10.1007/978-3-030-32962-4_86. URL: https://link.springer.com/chapter/10.1007/978-3-030-32962-4_86.

Dolphyne, F. (1988). *The Volta-Comoe Languages*. Ed. by L. Gerhardt. St. Augustin: Wissenschaftsverlag.

GhanaNLP (2021). *Khaya Translator Web App*. https://ghananlp.org/project/translatorwebapp. Accessed: January 17, 2024.

Gyasi, F. and Schlippe, T. (2023). "Twi Machine Translation". In: *Big Data and Cognitive Computing* 7.2, p. 114.

Kropp Dakubu, M. (2011). "The Central Togo Remnant Languages". In: *The Sociolinguistics of Ghana. The Frankfurt Contribution*. Ed. by K. Lotsu. Amsterdam: John Benjamins Publishing Company, pp. 141–194.

Lewis, M., Simons, G., and Fennig, C. (2015). *Ethnologue: Languages of the World*. 18th ed. Accessed: 4 May 2024. Dallas, Texas: SIL International. URL: https://www.ethnologue.com.

Musah, A. (2019). "The Challenges of Building Machine Translation Systems for Ghanaian Languages". In: *Journal of Computer Science and Applications* 7.1, pp. 1–8. URL: https://www.researchgate.net/publication/333801693.

Nekoto, W., Marivate, V., Matsila, T., Fasubaa, T., Kolawole, T., Fagbohunge, T., Adesanya, S., Alo, P., Adeyemi, S., Mokgesi-Mokgese, K., Sharifaith, S., Tada, S., Karani, S., and Bathseba, M. (2020). "Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 2144–2160. URL: https://aclanthology.org/2020.findings-emnlp.195.pdf.

Orife, I., Kreutzer, J., Sibanda, T., Whitenack, D., Simiyu, M., Martinus, L., Akera, J., Kinuthia, J., and Abade, S. (2020). "Leveraging Language Technology for Under-Resourced Languages: A Case Study of Igbo". In: *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 6855–6863. URL: https://aclanthology.org/2020.lrec-1.847.pdf.

Osam, E. (2003). "An Introduction to the Verbal and Multi-Verbal System of Akan". In: *Proceedings of the Workshop on Multi-Verb Constructions. Trondheim Summer School*. Accessed: 4 May 2024. URL: http://www.ling.hf.ntnu.no/tross/trondheimtravaux.pdf.

Osam, E., Duah, R., and Blay, A. (2011). "The Ga Language". In: *The Sociolinguistics of Ghana. The Frankfurt Contribution*. Ed. by K. Lotsu. Amsterdam: John Benjamins Publishing Company, pp. 195–238.

Salawu, S. (2021). "Low-Resource Machine Translation for African Languages". In: *Proceedings of the 2nd Workshop on Multi-disciplinary Approaches to Code-Switching*, pp. 47–56. URL: https://aclanthology.org/2021.multidisciplinary-codeswitching.7.pdf.

Siminyu, N., Mureithi, A., and Omondi, B. (2019). "Statistical Machine Translation for Kenyan Languages: Overcoming Data Scarcity and Complexity". In: *Proceedings of the African Conference on Computational Linguistics (AfriCL)*, pp. 45–54.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). "Attention is all you need". In: *Advances in neural information processing systems* 30.

Wilks, I. (1993). *Forests of Gold: Essays on the Akan and the Kingdom of Asante*. Athens: Ohio University Press.

Yamoah, K. (2023). "Efforts to Apply Natural Language Processing Technologies in Africa". In: *Science Engineering Entrepreneurship Design (SEED) Journal* 2.1.