Neural Machine Translation of Sinhala to English

D. G. P. Hansadi
(Reg. No.: MS23002142)

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2024

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Dr. Dılshan De Silva

Approved for MSc. Research Project:

MSc. Programme Co-ordinator, SLIIT

Approved for MSc:

Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Sign:  .....

D.G.P. Hansadi

Date: ……16/01/2025……..

ABSTRACT

Neural Machine Translation of Sinhala to English

D. G. P. Hansadi

MSc. in Information Technology

Supervisor: Dr. Dilshan De Silva

December 2024

This addresses the challenges of translating Sinhala in the form of Singlish, a unique and informal variation of Sinhala written in English letters, into standard English, aiming to bridge the communication gap for Sinhala-speaking individuals. The research explores the application of advanced transformer architectures such as T5, mBART, and MarianMT, each chosen for their proven capabilities in handling multilingual and low-resource language tasks. The study involves training and fine-tuning these models on a curated Singlish and English parallel sentences including dataset and evaluates their performance across several key metrics, including Quantitative and Qualitative analysis. The system developed as part of this research integrates practical features like speech-to-text for Singlish input, text-to-speech for English output, and translation history for user convenience, making it a comprehensive tool for real-time translation needs. The thesis not only aims to produce an effective translation system but also contributes valuable insights into optimizing transformer models for low-resource languages, offering benchmarks and techniques for future research in neural machine translation. By combining cutting-edge technology with a user-centric design, this research seeks to enhance accessibility for Sinhala speakers and sets the stage for advancing machine translation for underrepresented languages and dialects.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my supervisor, Dr. Dilshan De Silva, for his exceptional guidance, unwavering support, and encouragement throughout the course of this research. I am truly grateful for his patience and for always being available to discuss ideas, provide direction, and offer constructive criticism. His mentorship has not only guided me academically but has also been a source of great inspiration, motivating me to persevere through challenges and refine my work to a higher standard.

TABLE OF CONTENTS

List of Figures

List of Tables