

Informatics Institute of Technology
In Collaboration with
University of Westminster, UK



**HYBRID BERT-LSTM:
A TRANSFORMER AND LSTM BASED HYBRID
SINHALA CHATBOT SYSTEM USING TEXT
CLASSIFICATION
APPROACHA**

A Dissertation by
Mr. Chryshel Aloka Fonseka
W1761339 / 2019405

Supervised by
Ms. Sachinthani Perera

May 2023

This document was submitted in partial fulfilment of the
BEng (Hons) in Software Engineering degree requirements at the
University of Westminster

© The copyright for this project and all its associated products
resides with Informatics Institute of Technology

Abstract

With recent developments in the domain of Natural Language Processing (NLP) most of the day-to-day activities that involve natural language processing have been automated, including tasks like writing scripts, writing emails, customer support services, language translation etc. But in languages like Sinhala, those novel threads and techniques have yet to be applied and adapted because of language ambiguity and the lack of resources. Developing Sinhala language processing can contribute to its future advancements in areas such as text generation, chatbot applications, text summarization, language translations and etc.

This research focused on improving the Sinhala NLP domain by introducing a novel machine learning architecture called the Transformers architecture to chatbot application systems. This will increase the accuracy of Sinhala language chatbot systems. The Novel architecture introducing from this research study is a hybrid approach, which SinBERT and a LSTM layer combine to improve the accuracy of the model