



INFORMATICS
INSTITUTE OF
TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

**siNews : A Domain-Specific Hybrid Approach for Summarizing
Low-Resource Sinhala News**

A Project Proposal by

Mr. Kevin Paris

Supervised by

Mrs. Malsha Fernando

Submitted in partial fulfilment of the requirements for the BEng (Hons) Software
Engineering degree at the University of Westminster.

April 2025

ABSTRACT

A large number of news articles are created and published every day, and users can't browse through all available news to seek their interested news information. There is a lack of standardized datasets for Sinhala, which makes it challenging to train and evaluate NLP models effectively.

A combination of extractive and abstractive summarization with domain-specific fine-tuned pre-trained model. The model is trained with a custom-made, domain-specifically categorized data set to get the characteristics of each news domain.

Subject Descriptors: Machine learning~Transformer based models~pre-trained models

Keywords: Machine Learning, Text Summarization, NLP, Pre-trained Models, Low-resource Language Summarization