



INFORMATICS
INSTITUTE OF
TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration With

UNIVERSITY OF WESTMINSTER

**Automated Answer-Agnostic Diverse Question Generation with
Self-Attention Architectures**

A dissertation by

Mr. Inusha Manawadu

w1790317 / 20191059

Supervised by

Mr. Kandasamy Yogendirakumar

April 2023

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

ABSTRACT

Automatic generation of questions is a challenging task in Natural Language Processing that has attracted a lot of attention in recent years. Creating a variety of question categories, such as open-ended, true/false, and multiple-choice questions, that cover different aspects of the paragraph's content remains a challenge. In addition, existing methods have a tendency to generate questions that are identical in structure and content, resulting in a lack of diversity that could impact the accuracy of comprehension assessment. In order to improve comprehension assessment, an innovative approach is required that can generate a wide variety of question types while promoting question diversity.

Transformers have shown outstanding results across multiple sequence-to-sequence tasks in natural language processing. In this thesis, the author proposes a novel approach that uses the power of transformers to generate multiple types of paragraph-level questions. To generate open-ended and true/false questions, the T5 transformer model is fine-tuned using two distinct datasets. For the generation of open-ended questions, T5 model is fine-tuned on the SQuAD dataset, which is frequently utilized for question-answering tasks and the T5 model is fine-tuned on the BoolQ dataset, which contains boolean questions derived from Wikipedia articles, to generate true/false questions. In addition, for creating multiple-choice questions, the word embeddings are combined with Sense2Vec to extract semantic relationships between keywords and potential answer options. Through these approaches, it was able to generate a variety of questions that accurately capture the paragraph's content and meaning.

To evaluate the effectiveness of the approach, a combination of well-known metrics for evaluating language generation tasks which includes BLEU, METEOR, ROUGE, and Cosine Similarity, were used. The results indicate that this approach can effectively generate a variety of question types with high quality. Specifically, it achieved ROUGE-1 score of 45.40 and ROUGE-2 score of 25.34, BLEU-1 score of 13.58, BLEU-2 score of 26.04, and BLEU-3 score of 18.7, METEOR score of 43.7, and Cosine Similarity score of 52.