

Identifying AI-generated and Human-written Answers in Sinhala Using a Deep Learning Approach

R.A.D.K. Ranathunga^{1*}, R.A.H.M. Rupasingha², B.T.G.S. Kumara³

¹*Dept. of Computing and Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka, radkranathunga@std.appsc.sab.ac.lk*

²*Dept of Economics and Statistics, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka, hmrupasingha@gmail.com*

³*Dept. of Data Science, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka, kumara@appsc.sab.ac.lk*

As the use of Artificial Intelligence (AI) to answer questions increases, concerns about cheating also emerge, especially in languages like Sinhala, where efficient detection methods are lacking. This makes it easy for AI to provide incorrect or plagiarized versions of answers, undermining the fairness and creativity of education. There are similar AI-generated content detectors for the English language; however, no solution exists for the Sinhala language yet. The significance of this research lies in the proposed method's focus on the Sinhala language, addressing a major drawback of utilizing currently available deep learning-based detection systems to distinguish between AI-generated and human-generated content. The study begins with data collection, involving 1,000 academic questions, human-written answers, and AI-generated answers. The text pre-processing encompasses some operations including stemming, stop word removal and tokenization. To assess the variations, feature extraction techniques such as; term frequency-inverse document frequency (TF-IDF), Global Vectors for Word (Glove), Word Embedding vectors (Word2Vec), and Document Embedding vectors (Doc2Vec) are used. Among these, TF-IDF is identified as the most effective. Numerical data is then analyzed with machine learning techniques such as ANN and LSTM. Thus, 5-fold and 10-fold cross-validation is used in order to derive more stable evaluations. Performance of the models shows that the proposed LSTM has a higher accuracy of 86% compared to the ANN accuracy 85%. LSTM also demonstrated better recall, precision, f-measure and error rates. This research also prevents the academic dishonesty compared with other sources and contributes to the construction of Sinhala language processing assets. This work paves the way for future research to explore AI detection methods in other low-resource languages.

Keywords: *AI-generated, classification, deep learning, human-written, Sinhala language*