

## **A machine learning-based evaluation of English-to-Sinhala translation: comparing Google Translate, large language models, and human translators**

Jayathilaka K.M.D.P.S.D.<sup>1\*</sup>, Rubasinghe T.D.<sup>1</sup>, and Kumara B.T.G.S.<sup>1</sup>

<sup>1</sup>Department of Software Engineering, Faculty of Computing,  
Sabaragamuwa University of Sri Lanka, Sri Lanka

\*kmdpsdhiyan@std.appsc.sab.ac.lk

Reliable translation from English to Sinhala is still a great challenge for many sophisticated translation systems using Sinhala as a low-resource language. Although Google Translate is widely used for translation purposes, recent breakthroughs in large language models such as ChatGPT and DeepSeek provide entirely new opportunities for translation tasks. This study proposes one of the first thorough comparative analyses of English-Sinhala translation systems compared with human translation, both qualitatively and quantitatively. Google Translate, ChatGPT, DeepSeek, and human translations done by native Sinhala speakers were compared for translation quality on a carefully prepared dataset of 150 English sentences for general, technical, and academic purposes. Translation quality was compared using BLEU, METEOR, and COMET scores, in addition to human assessment of fluency, grammatical accuracy, and semantic translation quality done by qualified human raters using a prepared rubric with inter-rater reliability tests. Machine learning models were also prepared for predicting translation quality using language-based predictors for translation efficiency and translation system identification. The experimental results show that human translations were rated highest on all translation quality measures. Among the automatic translation systems, LLM-based translation systems performed better on contextual understanding of complex sentences than Google Translate, which performed reasonably on simple inputs. Correlation tests show that COMET correlates better with human translation quality than BLEU and METEOR. Moreover, the prepared machine learning models were able to detect translation quality trends accurately for translation system predictions, making these models promising for translation quality assessment in low-resource language environments.

**Keywords:** *Large language models; Low-resource languages; Machine translation; Sinhala language; Translation evaluation*