



## Deep Learning based Sinhala Virtual Assistant

L.R.L.D. Siwurathna, H.K.I.R. Kalhara, M.D.A. Perera, K.L.K. Sudheera, W.G.C.A. Sankalpa, and J.C.S Kadupitiya

*Department of Electrical and Information Engineering, Faculty of Engineering, University of Ruhuna, Sri Lanka.*

*<sup>b</sup> Microsoft Corporation, USA.*

### Abstract

Virtual Assistants have been at the forefront of end-user engagement for a long time. The primary intention of a virtual assistant is to give a unique user-friendly interface that allows the user to work more quickly and comfortably. NLP is the widely used technique to meet the advanced criteria and provide the capability that the user expects from a virtual assistant. Even though a few virtual assistants have previously been developed, they primarily respond to the English language and frequently fail to distinguish accents other than the usual English accents, American or British; and do not cater to the sentiments of the Sri Lankan service sector. We have taken the initiative to develop a virtual assistant that approaches interaction between the user and the machine using the Sinhala language, utilizing Deep Learning and NLP. The system accepts and responds to queries in either text or voice by the user. The input query is fed to the NLP component and goes through the required pre-processing procedures and passes to the tokenizer. A Sinhala-specific tokenizer was designed specifically for this model, which can be considered another accomplishment in the attempt to achieve the main objective. The deep learning component generates a specific vector for the query after feature engineering together with the attention mechanism. At the decoder of the transformer, it generates the most probable next word and ultimately generates the final output query. The performance of the purposed system was evaluated using the Turing Test, which is a measure of how well a machine can imitate human behaviour, with an unbiased sample group that had not seen the model previously. The test received good results where a 7.05 average rating out of 10 was recorded. Subject to the availability of sufficient training data, the proposed framework can be implemented in multiple service domains.

Keywords: *Deep Learning, Natural Language Processing, Question Answering System, Sinhala Virtual Assistant, Transformers.*

---

Corresponding Author: *kushan@eie.ruh.ac.lk*