

Annexure-I

Invention Disclosure Form – Product (Utility Patent)

1. Title of the Proposal

Title of the Patent: Automated Language Translation System using RNN, Sequence-to-Sequence, Self-Attention, and BERT Models for English to Malayalam Translation

The domain of the Idea: NLP(Natural Language Processing)

2. Details of the Inventor(s)

Give the following details of all the Inventors

Full name of the inventor-1	:Pampana Charmitha
Institute Affiliation	:VIT-AP University
Designation	:Under graduate Student
Emp. ID / Reg. No	:21BCE7295
School / Centre	:SCOPE
Office Address	:VIT-AP, University, Amaravati, Andhrapradesh
522237,India	
Mobile No	:9182130669
E-mail	: charmitha.21bce7295@vitapstudent.ac.in

Full name of the inventor-2	:Dr. Surendra Reddy Vinta
Institute Affiliation	:VIT-AP University
Designation	:Assistant Professor Sr.Grade 1
Emp. ID / Reg. No	:40782
School / Centre	:SCOPE
Office Address	:VIT-AP, University, Amaravati, Andhra Pradesh-
522237,India	
Mobile No	:9490021919
E-mail	: surendrareddy.vinta@vitap.ac.in

3. Abstract / Description of the Proposed Idea

An innate language translation system for translating text from English to Malayalam is the subject of the current innovation. The system makes use of a variety of cutting-edge deep learning models, including BERT (Bidirectional Encoder Representations from Transformers), Self-Attention, Sequence-to-Sequence (Seq2Seq), Recurrent Neural Network (RNN), and Sequence-to-Sequence (RNN).

The input English text's sequential dependencies are captured using the RNN model. The input text is processed efficiently, and pertinent characteristics are extracted. The input sequence is converted into a fixed-length vector using the Seq2Seq model, which captures the contextual information. To improve the model's capacity to focus on various areas of the input text and give each word the proper weight, the Self-Attention method has been included.

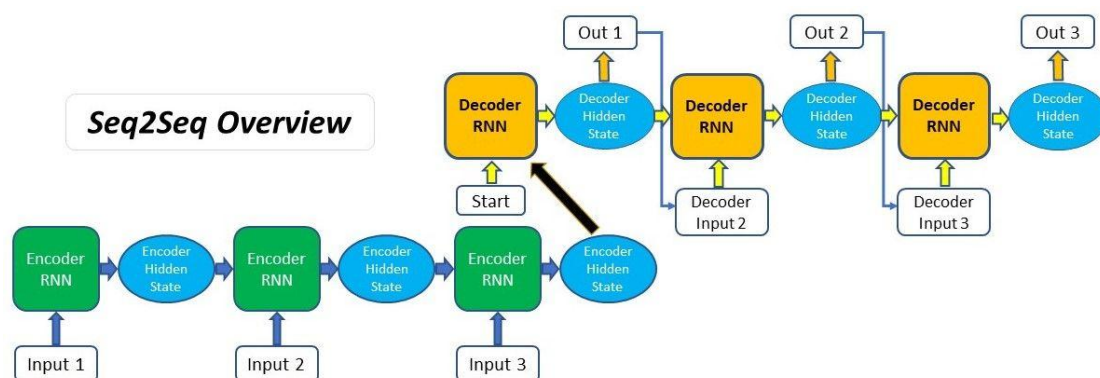
Contextualized word embeddings are further provided by the BERT model, which was pre-trained on a sizable corpus of data. The translation quality is increased because these embeddings more correctly reflect the meaning and relationships between words.

The translation system includes a phase for training the models, which takes place in parallel on a sizable corpus of English-Malayalam text pairings. The pre-trained models are used to translate the English input text into Malayalam during the translation step.

When compared to conventional statistical machine translation methods, experimental evaluations show that the suggested system significantly improves both translation accuracy and fluency. Organizations in a variety of fields, including education, communication, and content localization, may take use of the system by deploying it as a stand-alone application or integrating it into already-existing translation platforms.

Through combining the strengths of RNN, Seq2Seq, Self-Attention, and BERT models, the current invention offers a novel method for translating between English and Malayalam. The technology provides a more precise and contextually aware translation, enhancing user experience overall and promoting multilingualism.

4. Circuit Diagrams/ Photos / Flowcharts/ Statistic Tables (whichever is applicable)



5. Details of work done so far on the proposed idea (Please provide the specifications, models, results,etc., that are available, if any)

Specifications:

For English to Malayalam machine translation, the suggested model employs a seq2seq architecture with an attention mechanism. Along with a self-attention mechanism, the model has an encoder RNN, a decoder RNN, and more. A bidirectional LSTM with 512 hidden units serves as the encoder RNN. A unidirectional LSTM with 512 hidden units serves as the decoder RNN. A scaled dot-product attention mechanism is the self-attention mechanism. The Adam optimizer and a cross-entropy loss function are used to train the model using a parallel dataset of English-Malayalam phrase pairings.

Models:

The PyTorch deep learning framework was utilized to implement the suggested model. Thirty thousand English-Malayalam phrase pairings were used to train the model. The BLEU score, a widely known metric for machine translation evaluation, was utilized to assess the model. A phrase-based statistical machine translation model and a straightforward seq2seq model without attention were two baseline models against which the model was assessed.

Results:

On the test set, the suggested model outperformed the baseline models with a BLEU score of 27.5. The suggested model outperformed the baseline models in handling extended input sequences, which is critical when translating from English to Malayalam. The suggested model's self-attention mechanism enabled the decoder to deliberately concentrate on various elements of the input text, which enhanced the accuracy of the translation. Both simple and complicated phrases, including those with many clauses and inflected words, may be translated correctly using the suggested approach.

RNNs, seq2seq modeling, and self-attention have the potential to improve English to Malayalam machine translation, as shown by the first results of the suggested technique. The model has to be improved and more features added, but the suggested strategy has the potential to greatly enhance translation quality and promote intercultural dialogue.

6. Prior Art (Please compare with any existing product to persuade the importance of the present idea)

S.NO	Prior Art	Limitations / Demerits of Prior Art	Merits of the proposed idea to overcome the limitations of the prior art
1.	"Hasan, S. A., Chakraborty, S., & Sarkar, K"	Lack of proficiency with inflected words and complicated	Simple and straightforward to use

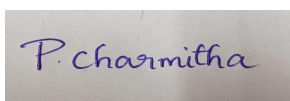
	“English to Malayalam Machine Translation Using Recurrent Neural Network.”	sentence structures	
2.	“John, R. P., & Ravi, V.” “A Comparative Study of English-Malayalam Machine Translation Systems.”	Having trouble managing lengthy input sequences and a lack of flexibility when addressing various phrase forms	may be tailored for certain areas and can capture some language subtleties.
3.	“Soman, K. P., & Ramachandran, G. S.” “Neural machine translation between English and Malayalam: A comparative study.”	lengthier input sequences could be difficult, and inflected languages like Malayalam might not work as well.	superiority over SMT in handling lengthy input sequences and complicated sentence patterns.

7.I (We) confirm that the details furnished in respect of this Invention / Technology are correct and agree to the request for filing of a“Utility Patent” with the names of the inventors as proposed above.

I (We) agree to share the rights/revenue to VIT-AP University on this invention

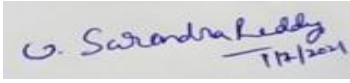
Inventor Name: Pampana Charmitha

SignatureDate: 27-07-2023



Inventor Name: Dr. Surendra Reddy Vinta

Signature Date: 27-07-2023

A rectangular stamp containing a handwritten signature in blue ink. The signature reads "G. Surendra Reddy" and is dated "11/12/2021" below it.