COMPARITIVE ANALYSIS OF GOOGLE TRANSLATOR AND AI4BHARAT TRANSLATOR

CLOUD COMPUTING

A translation technology study

Course Instructor : Dr.Animesh Chaturvedi

UNDER GUIDANCE OF

Dr . Animesh Chaturvedi
Researcher

OUR TEAM

Ch.Srinivasa Sai
21BDS012

K.Abhiram
21BDS029

R . Vinay Kumar
21BDS056

OVERVIEW OF GOOGLE TRANSLATOR AND AI4BHARAT TRANSLATOR

- Both Google Translate and AI4Bharat are significant players in the field of machine translation, but they have different focuses and strengths
- Focus of google translator is global translation across a vast number of languages (over 100)
- Focus of ai4bharat translator is on Indian languages, they are 22 scheduled languages, including multiple scripts for low-resource languages)

🔗 **GOOGLE TRANSLATOR**   🔗 **AI4 BHARAT TRANSLATOR**

PROBLEM STATEMENT

In light of Google Translator's inefficiencies with low-resource and certain Indian languages, can we find more effective alternatives to Google Translator? How do these alternatives compare in terms of accuracy and contextual meaning, aiming to deliver superior translations in diverse linguistic contexts?

MOTIVATION

Encountering challenges with Google Translator's inefficiencies in low-resource languages, the motivation behind opting for AI4Bharat Translator is twofold. Not only does it specialize in Indian languages, but it also addresses the pressing issue of accurately translating low-resource languages, promising a more comprehensive and effective solution for cross-lingual communication.

TRAINING METHODS OF GOOGLE TRANSLATOR

NEURAL MACHINE TRANSLATION (NMT) :

- The core technology behind Google Translate is NMT.

- NMT uses artificial neural networks trained on massive amounts of bilingual text data.

STATISTICAL MACHINE TRANSLATION (SMT) :

- While NMT is the dominant approach, SMT still plays a role in Google Translate.

- SMT uses statistical analysis of large parallel text data to identify patterns and translate text based on these patterns.

NMT VS SMT

Feature	Neural Machine Translation	Statistical Machine Translation
Mechanism	It translates sentence by sentence.	It translates word by word or phrase by phrase.
Underlying Model	Utilizes attentional decoder-encoder networks, often deep learning models like Recurrent Neural Networks (RNNs) or Transformers.	Relies on statistical models based on linguistic and probabilistic principles. Commonly uses phrase-based or word-based models.
Multilingual/Multi-domaintranslation	NMT excels in seamlessly handling multilingual translation and adapting to diverse domains through shared parameters and contextual learning.	SMT, on the other hand, may face challenges in efficiently managing multiple languages and domains, often requiring separate models and significant retraining.
Vocabulary/Rare word Problem	NMT models, with continuous embeddings, tend to handle rare words better, capturing nuanced meanings effectively.	SMT systems may struggle with rare words, relying on predefined vocabularies and facing challenges in accurately translating less common terms.

Translation Done using Google Translator :

Hindi	English	Eng_Trans
ऐ मेरे बन्दों! आज न तुम्हें कोई भय है और न तुम शोकाकुल होगे।	O My servants, today no fear is on you, neither do you sorrow	O my prisoners! Today you have no fear nor you will be heartbroken
वही है जो आकाशों में भी पूज्य है और धरती में भी पूज्य है और वह तत्वदर्शी, सर्वज्ञ है	And it is He who in heaven is God and in earth is God; He is the All - wise, the All - knowing.	He is the one who is also revered in the sky and is also revered in the earth and he is a philosopher, omnisc.
रहे वे लोग जिन्होंने इनकार किया, तो उनके लिए तबाही है। और उनके कर्मों को अल्लाह ने अकारथ कर दिया	But as for the unbelievers, ill chance shall befall them! He will send their works astray.	Those who refused, so there is havoc for them.And Allah made their deeds unconscious
फिर कैसी रही मेरी यातना और मेरे डरावे?	How then were My chastisement and My warnings?	Then how was my torture and my scared?

TRAINING METHODS OF AI4 BHARAT

Transformer-based Multilingual NMT:

- This model uses the powerful Transformer architecture, specifically tailored for Indic languages.
- It allows for parallelized processing and attention mechanisms, leading to faster training and better performance.

Multilingual Training :

- The model is trained on a massive dataset of parallel text across all 22 scheduled Indic languages and English.
- This enables it to learn the relationships between different languages and generate accurate translations.

Script Unification:

- To leverage lexical sharing between languages, IndicTrans2 adopts script unification wherever feasible.
- This means that languages like Kashmiri, Manipuri, and Sindhi, which use different scripts, are converted to a common script like Devanagari before translation and then converted back after translation.

TRANSFORMER ARCHITECTURE
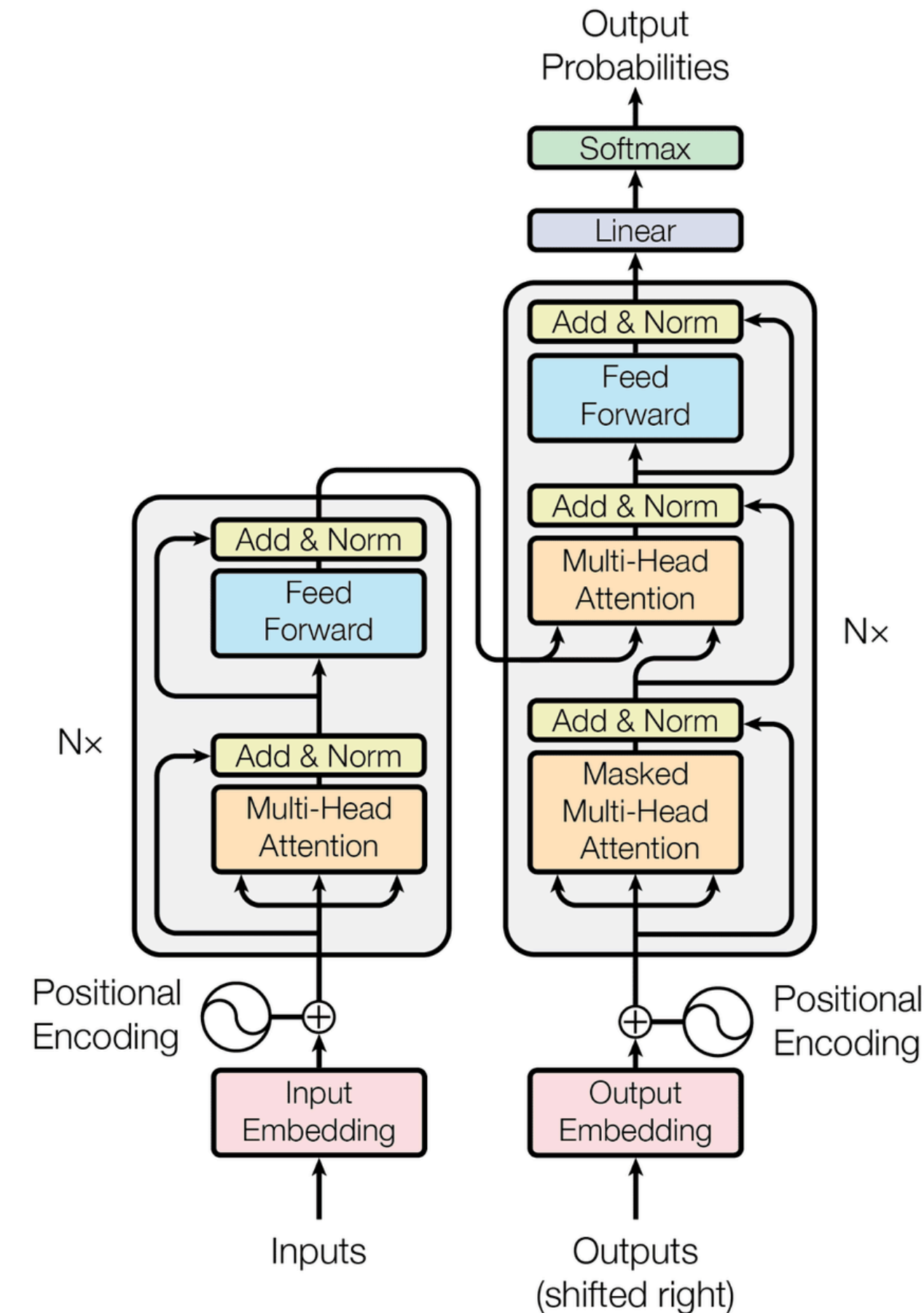
Translation done using ai4bharat Translator :

| Hindi	English	AI4_Bharat_trans
ऐ मेरे बन्दों! आज न तुम्हें कोई भय है और न तुम शोकाकुल होगे।	O My servants, today no fear is on you, neither do you sorrow	O My servants! Today you will have no fear nor will you grieve
वही है जो आकाशों में भी पूज्य है और धरती में भी पूज्य है और वह तत्वदर्शी, सर्वज्ञ है	And it is He who in heaven is God and in earth is God; He is the All - wise, the All - knowing.	He is the one who is God in the heavens and God on the earth. He is the Wise and the All Knowing.
रहे वे लोग जिन्होंने इनकार किया, तो उनके लिए तबाही है। और उनके कर्मों को अल्लाह ने अकारथ कर दिया	But as for the unbelievers, ill chance shall befall them! He will send their works astray.	But as for those who disbelieve, for them is perdition. And He will waste their deeds.
फिर कैसी रही मेरी यातना और मेरे डरावे?	How then were My chastisement and My warnings?	How [terrible] were My punishment and My warnings!

GOOGLE TRANSLATOR (NMT) ARCHITECTURE

ENCODER-DECODER ARCHITECTURE

- The fundamental architecture common to both Neural Machine Translation (NMT) and Statistical Machine Translation (SMT) models involves an encoder that compresses the source language text, and a decoder that utilizes this representation to generate the translated text in the target language.

ATTENTION MECHANISM

- Attention mechanism, a crucial component of NMT models, enhances translation accuracy and fluency by enabling the decoder to focus on specific parts of the encoded source language text during the generation of translations.
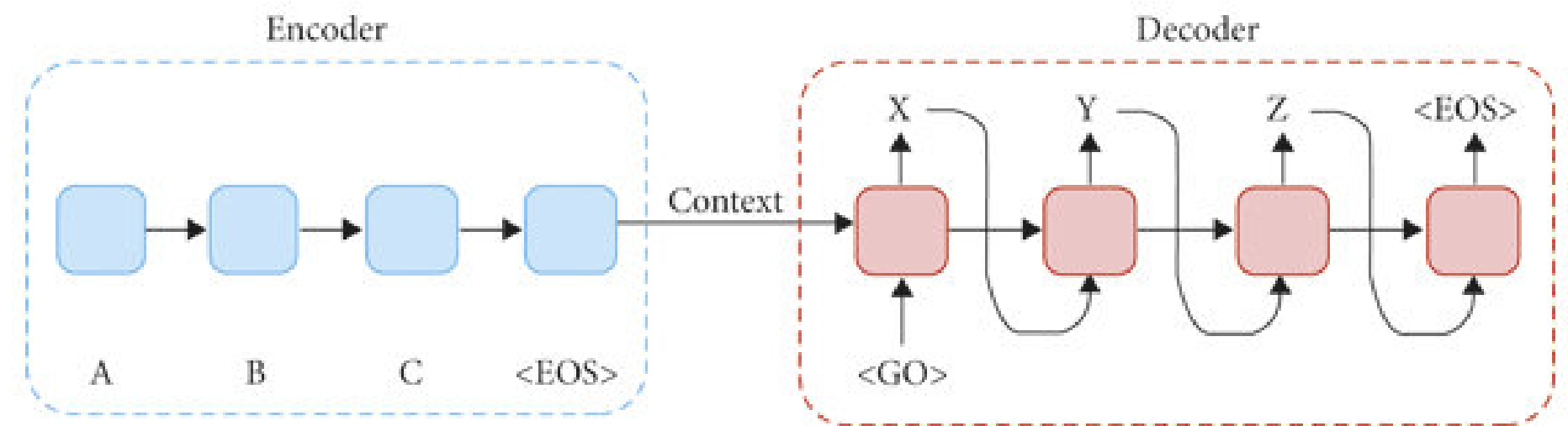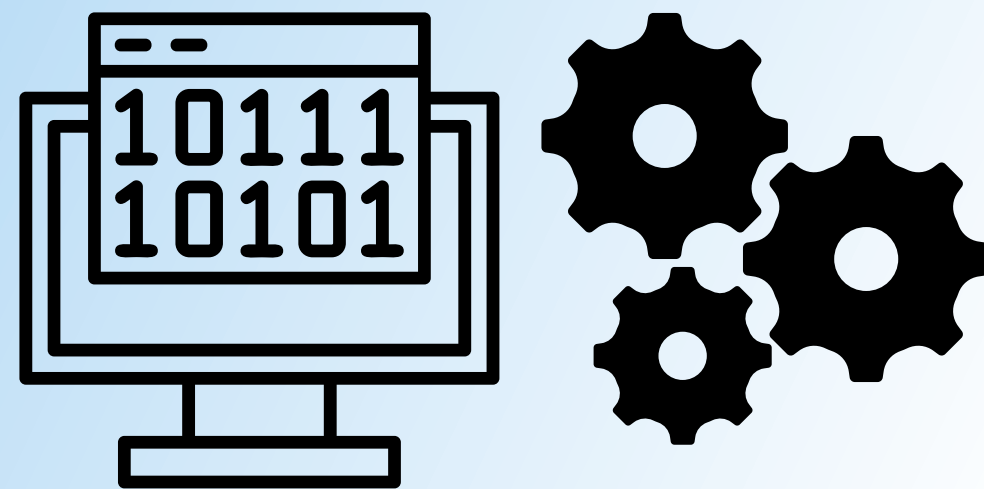
MULTI-LAYER PERCEPTRON

- Used in both NMT and SMT models to process the encoded representation and generate the translated text.

LONG SHORT-TERM MEMORY (LSTM) NETWORKS

- These are a type of recurrent neural network used in some NMT models to capture long-range dependencies in the source language text, which is important for accurate translations of complex sentences.
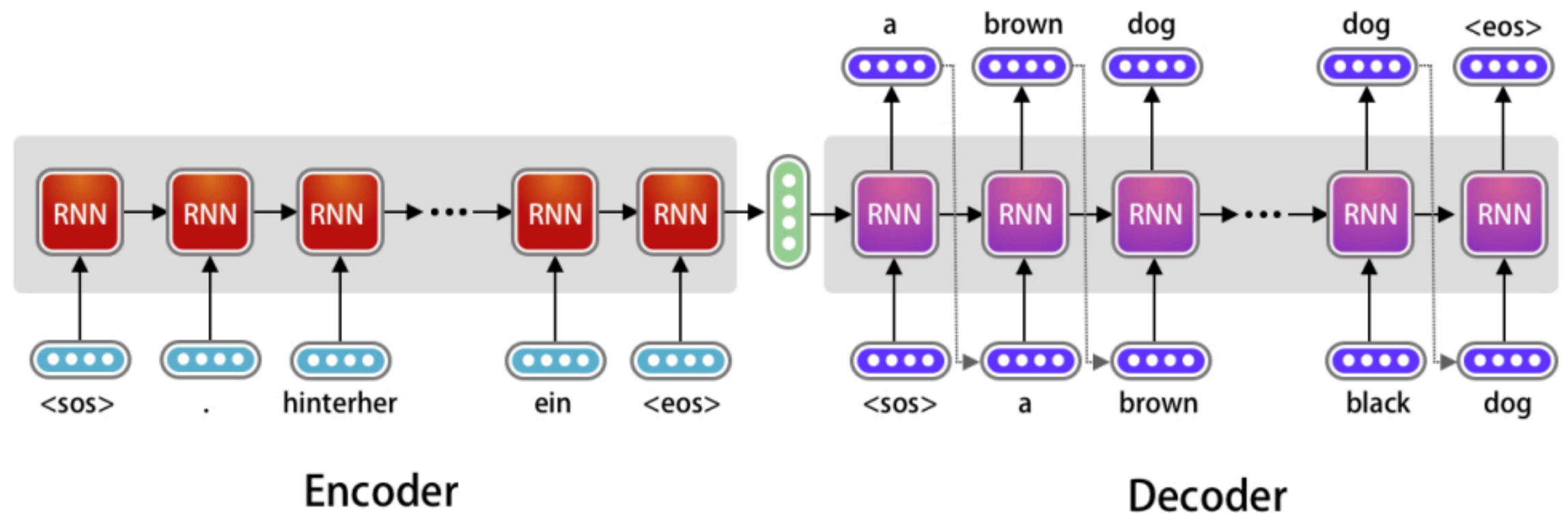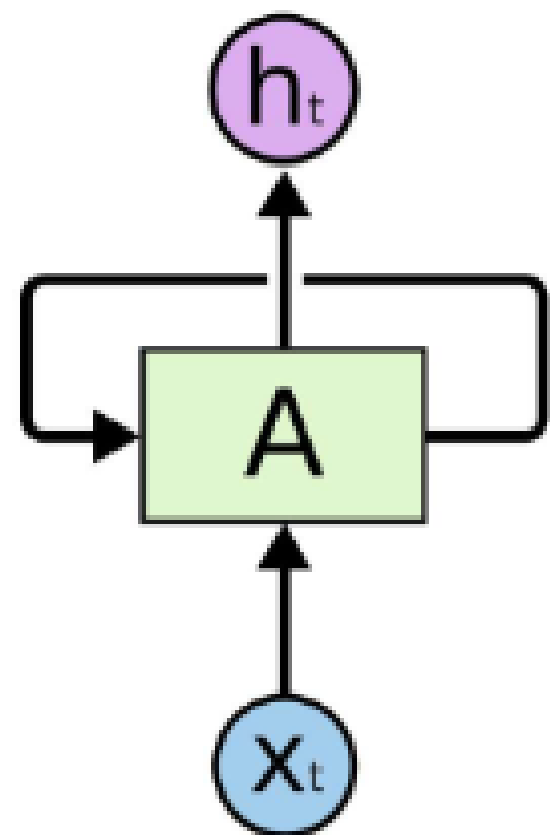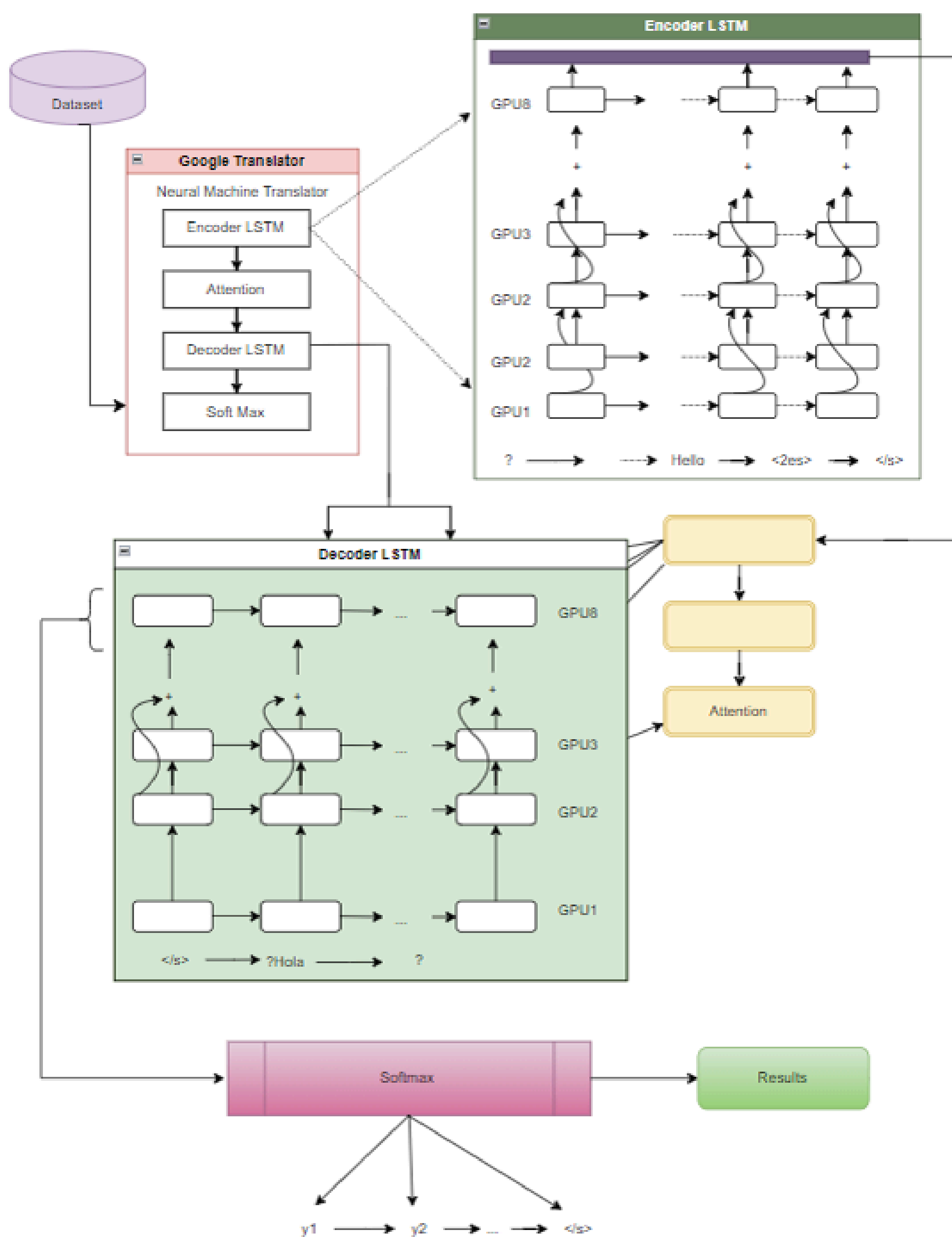
NMT COMPONENTS

ENCODER
- EMBEDDING LAYER
- RNN CELLS

DECODER
- RNN CELLS
- EMBEDDING LAYER

# INDICTRANS2 NEURAL TRANSLATION ARCHITECTURE

ENCODER-DECODER ARCHITECTURE

- IndicTrans2, akin to Google Translate, utilizes an encoder-decoder architecture where the encoder compresses the source language text into a hidden representation, and the decoder subsequently employs this representation to produce the translated text in the target language.

MULTI-HEAD ATTENTION

- This mechanism allows the decoder to attend to different parts of the encoded source language text, leading to more context-aware translations.

POSITIONAL ENCODING

- This technique helps the model understand the order of words in the sentence, which is crucial for accurate translations of languages with complex syntax like Hindi and Urdu.
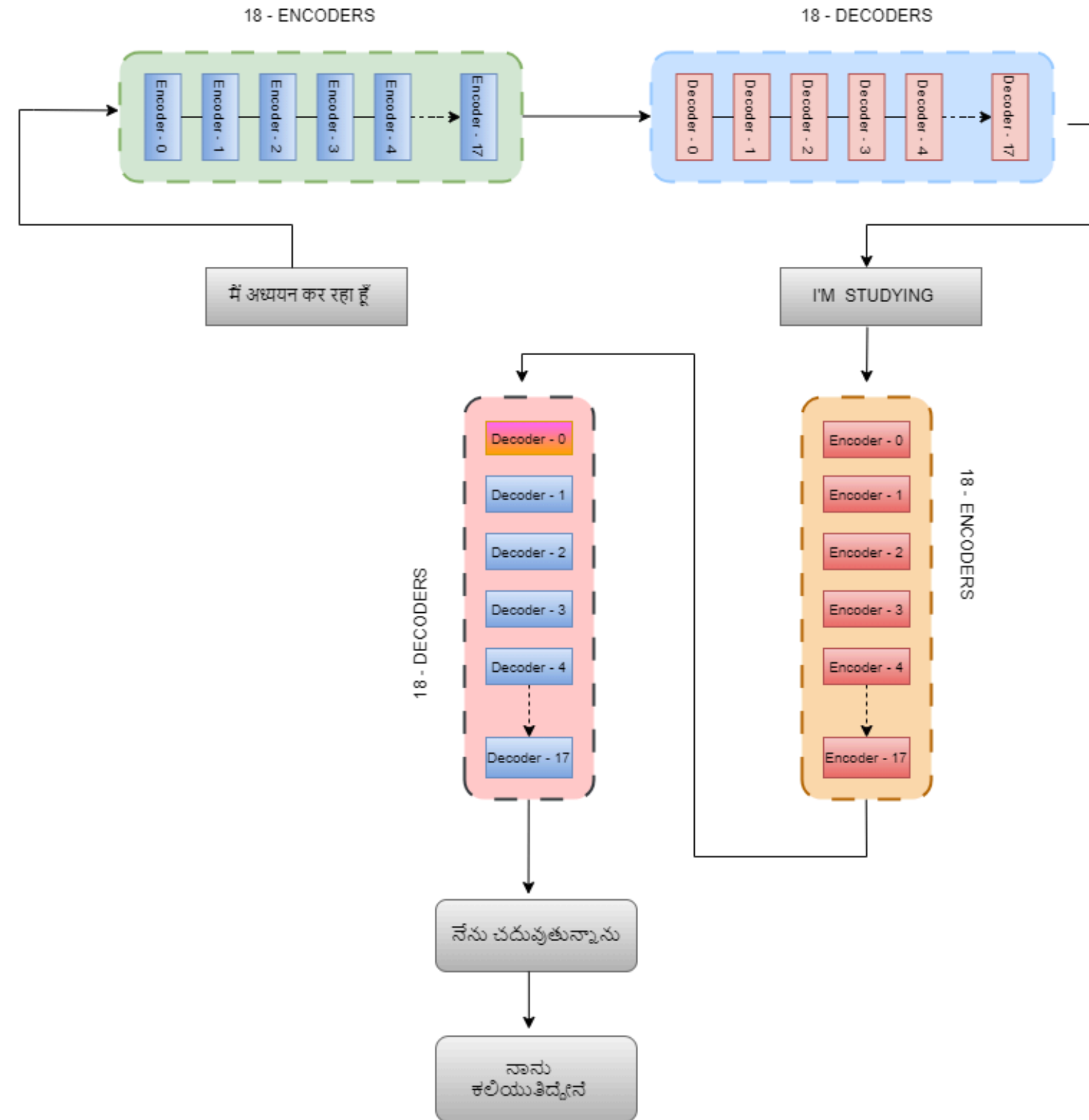
COMPARISON METRICS

Semantic Textual Similarity (STS)

1. Evaluates the semantic equivalence between two texts, focusing on capturing the underlying meaning and context. It measures how similar or equivalent two texts are in meaning, beyond simple word overlap, making it suitable for tasks like paraphrase identification and text summarization.
2. Relies on deep learning techniques trained on large datasets with annotated similarity scores. These models capture nuanced semantic relationships effectively, making them applicable across various domains and languages.
3. Primarily used for tasks requiring an understanding of semantic equivalence between texts, such as duplicate detection, sentiment analysis, and paraphrase identification.

BiLingual Evaluation Understudy (BLEU)

1. Assess the quality of machine-translated text by comparing it to human reference translations. It measures translation fluency and adequacy by examining the overlap of n-grams between the machine-generated and reference translations, primarily focusing on linguistic precision and adequacy.
2. Does not require deep learning techniques for computation. It is computed based on simple n-gram overlap between the machine-generated and reference translations. While it may not capture semantic nuances, it provides a quick and efficient metric for evaluating translation quality.
3. Mainly used for evaluating the quality of machine translation systems. It provides a standardized metric for comparing different translation models and assessing their performance across various languages and domains.

COMPARISON METRICS

Semantic Textual Similarity (STS) & BiLingual Evaluation Understudy (BLEU)

- **0: No similarity. This suggests that the machine translation bears no resemblance to the reference translations.**
- **0.5: Moderate level of similarity. The machine translation partially captures the meaning of the reference translations but may have significant errors or omissions.**
- **1: Perfect similarity. The machine translation is identical to the reference translations, conveying the same meaning accurately.**
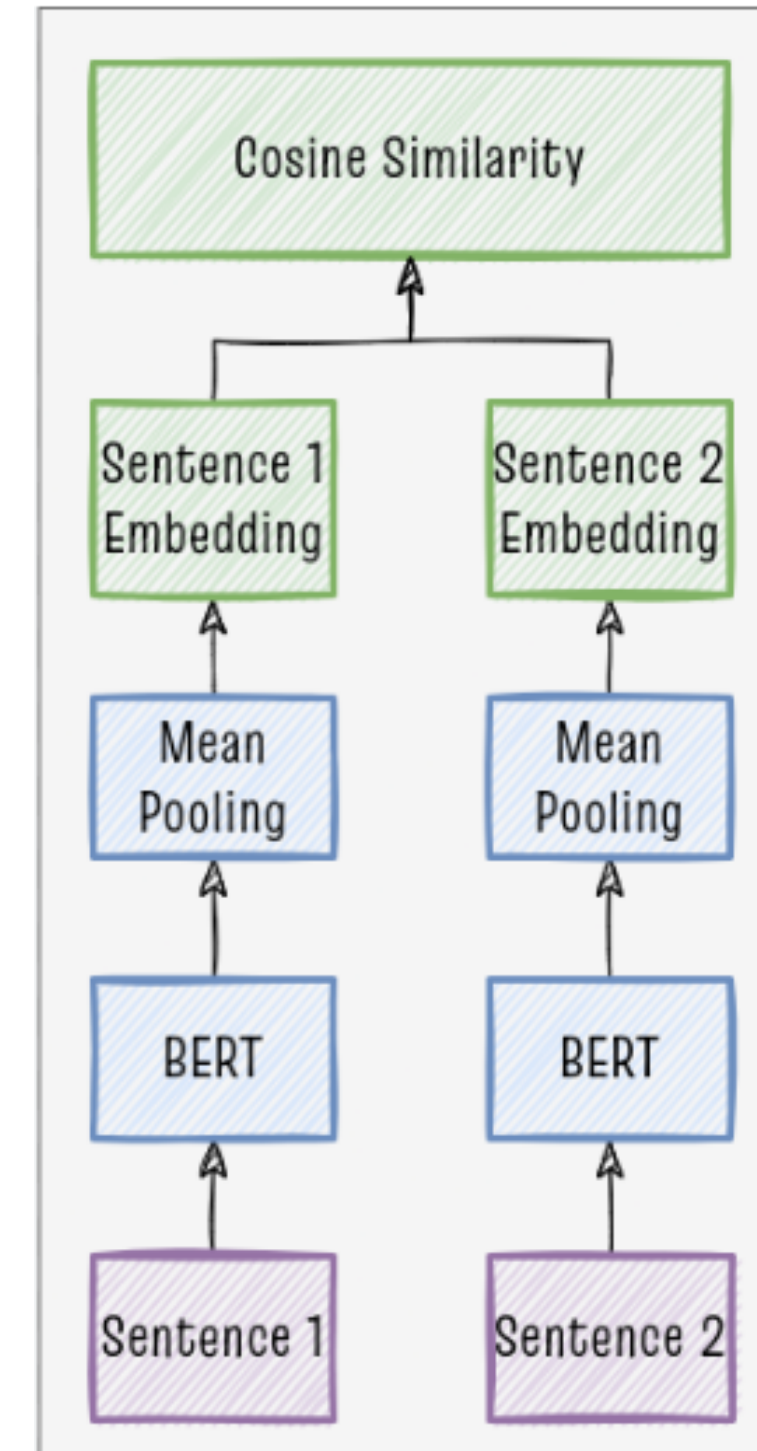
FINAL RESULTS :

English	Eng_Trans	AI4_Bharat_trans	STS_for_google_trans	STS_for_AI4bharat	Blue_Score_google_trans	Blue_Score_AI4bharat
O My servants, today no fear is on you, neither do you sorrow	O my prisoners! Today you have no fear nor you will be heartbroken	O My servants! Today you will have no fear nor will you grieve	0.2912	0.5101	0.002569	0.015894
And it is He who in heaven is God and in earth is God; He is the All - wise, the All - knowing.	He is the one who is also revered in the sky and is also revered in the earth and he is a philosopher, omnisc.	He is the one who is God in the heavens and God on the earth. He is the Wise and the All Knowing.	0.0531	0.9191	0.005474	0.009825
But as for the unbelievers, ill chance shall befall them! He will send their works astray.	Those who refused, so there is havoc for them.And Allah made their deeds unconscious	But as for those who disbelieve, for them is perdition. And He will waste their deeds.	0	0.7391	0	0.056637
How then were My chastisement and My warnings?	Then how was my torture and my scared?	How [terrible] were My punishment and My warnings!	0	0.2606	0	0.003874

Indian Institute of Information Technology
Dharwad

THANK YOU!

TEAM PROFILES

CH . SRINIVAS SAI   K . ABHIRAM   R . VINAY KUMAR