

# **Computational Linguistics for Indian Languages ( CS689A )**

## **ASSIGNMENT 3**

**SUBMITTED BY: PRASHIK GANER**

**ROLL NO: 231110037**

# Question 1

---

## **Translation Model:**

Translation models in natural language processing (NLP) aim to automatically convert text from one language to another. They can be broadly categorized into statistical and neural machine translation (NMT) models.

## **NLLB(No Language Left Behind):**

NLLB is a multilingual neural machine translation model developed by Facebook AI. It aims to bridge language barriers by providing high-quality translations across multiple languages.

## **IndicTrans/IndicIndic:**

IndicTrans is a specialized translation model developed specifically for Indian languages, addressing the unique linguistic characteristics and challenges of the Indian subcontinent.

Refer [cl-assignment-3-indictrans.ipynb](#), [cl-assignment-3-nllb.ipynb](#)

# QUESTION 2

## BLUE Score:

BLEU (Bilingual Evaluation Understudy) is a metric for evaluating the quality of machine-generated translations by comparing them to one or more reference translations. It measures the n-gram overlap between the candidate translation and reference translations.

## ROUGE SCORE:

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is a set of metrics for evaluating summarization and machine translation outputs. It measures the overlap of n-grams and word sequences between the candidate and reference texts.

## Evaluation Report: Translation Models Comparison:

### NLLB Model:

Translation Direction	BLEU Score	ROUGE-1 (R)	ROUGE-1 (P)	ROUGE-1 (F)	ROUGE-2 (R)	ROUGE-2 (P)	ROUGE-2 (F)	ROUGE-L (R)	ROUGE-L (P)	ROUGE-L (F)
English to Hindi	0.6388	0.5741	0.6078	0.5851	0.3204	0.3382	0.3254	0.5120	0.5430	0.5224
Hindi to English	0.6533	0.5858	0.6104	0.5939	0.3673	0.3804	0.3700	0.5579	0.5794	0.5647
Hindi to Marathi	0.6888	0.5734	0.6325	0.5970	0.3311	0.3663	0.3452	0.5546	0.6105	0.5769
Marathi to Hindi	0.5749	0.5082	0.5516	0.5239	0.2491	0.2684	0.2552	0.4639	0.5008	0.4772

# IndicTrans Model:

Translation Direction	BLEU Score	ROUGE-1 (R)	ROUGE-1 (P)	ROUGE-1 (F)	ROUGE-2 (R)	ROUGE-2 (P)	ROUGE-2 (F)	ROUGE-L (R)	ROUGE-L (P)	ROUGE-L (F)
English to Hindi	0.6872	0.6242	0.6164	0.6150	0.3927	0.3846	0.3849	0.5701	0.5626	0.5618
Hindi to English	0.7239	0.6244	0.6299	0.6243	0.4207	0.4224	0.4192	0.5953	0.5984	0.5941
Hindi to Marathi	0.5956	0.3870	0.3946	0.3873	0.1394	0.1405	0.1388	0.3538	0.3592	0.3534
Marathi to Hindi	0.6039	0.5276	0.5326	0.5251	0.2686	0.2725	0.2673	0.4733	0.4760	0.4704

# ChatGPT Model:

Translation Direction	BLEU Score	ROUG E-1 (R)	ROUGE-1 (P)	ROUGE-1 (F)	ROUGE-2 (R)	ROUGE-2 (P)	ROUGE-2 (F)	ROUGE-L (R)	ROUGE-L (P)	ROUGE-L (F)
English to Hindi	0.6161	0.5296	0.5347	0.5277	0.2700	0.2747	0.2700	0.4726	0.4801	0.4725
Hindi to English	0.7570	0.6853	0.6696	0.6737	0.4964	0.4810	0.4848	0.6618	0.6436	0.6490
Hindi to Marathi	0.4838	0.2835	0.2866	0.2831	0.0782	0.0778	0.0775	0.2618	0.2640	0.2611
Marathi to Hindi	0.6618	0.6062	0.6280	0.6138	0.4219	0.4390	0.4281	0.5761	0.5955	0.5828

Refer cl-assignment-3-question-2.ipynb

# QUESTION 3

---

## Observations:

1. The IndicTrans model generally outperforms the other two models across most translation directions in terms of both BLEU and ROUGE scores.
2. ChatGPT shows particularly strong performance in Hindi to English translation.
3. NLLB performs relatively better in Hindi to Marathi translation compared to the other models.

BLUE scores of every model turned out to be better than the ROUGE scores individually. NLLB stands out for its accuracy and fluency in translating from Hindi to English, with high scores. IndicTrans also performed well, especially in translating from English to Hindi and Hindi to English. However, ChatGPT, while decent, falls short compared to NLLB and IndicTrans in some aspects.

IndicTrans is really good at translating between languages like Hindi and Marathi, as we can see from its high scores. NLLB also does a decent job with Indic languages, getting pretty good scores too. ChatGPT might need some work for Indic languages, but both NLLB and IndicTrans are great for translating within the Indic language group.

But for Marathi to Hindi, CHATGPT gave the best scores as compared to NLLB, IndicTrans/InndicIndic.