

### QUESTION 3:

#### BLUE Scores

	Hi to En	En to Hi	Hi to Gu	Gu to Hi
NLLB	0.718543	0.755346	0.763142	0.758607
IndicTrans2	0.713449	0.750331	0.755517	0.750858
ChatGPT	0.702682	0.753073	0.750935	0.740908

#### ROUGE Scores(F1 scores):

	Hi to En	En to Hi	Hi to Gu	Gu to Hi
NLLB	0.6016	0.5560	0.5232	0.5744
IndicTrans2	0.6477	0.5797	0.5322	0.5550
ChatGPT	0.5601	0.5904	0.4037	0.4436

#### BLUE (Bilingual Evaluation Understudy) Score:

- BLUE is a metric used to evaluate the quality of machine-translated text by comparing it to reference translations.
- It measures the precision of the machine translation by calculating the geometric mean of n-gram overlaps between the candidate translation and the reference translation.
- Higher BLUE scores indicate better translation quality, with a maximum score of 1 representing perfect translation.

## **ROUGE (Recall-Oriented Understudy for Gisting Evaluation) Score:**

- Rouge measures the recall of the system-generated summaries or translations by comparing them to one or more reference summaries or translations.
- ROUGE scores are calculated based on various factors such as n-gram overlap, word overlap, and longest common subsequence.

## **OBSERVATIONS:**

### **IndicTrans2:**

1. Performs well overall, with competitive scores across all language pairs.
2. Achieves the highest ROUGE F1 scores for translations from Hindi to English.
3. Demonstrates relatively consistent performance across translation in terms of ROUGE F1 scores.
4. Scores slightly lower in BLUE scores compared to NLLB.

### **NLLB:**

1. Achieves the highest ROUGE F1 score for translations from Gujarati to Hindi.
2. Performs well in most translation directions, with relatively consistent ROUGE F1 scores.
3. Attains the highest BLUE scores across all translation directions.

### **ChatGPT:**

1. Perform good but lags behind the other two models in terms of ROUGE and BLUE scores.
2. Not good in translating hindi to gujrati or gujarati to hindi.