# Assignment - 3

## Computational Linguistics for Indian Language: CS689

Govind Sharma(231110015)

This assignment taught me a lot of things. I learned how to evaluate models like NLLB-200, IndicTrans, chatGPT, how to handle datasets, and work with languages that have limited resources. Here's what I learned:

## Observations and Learnings:-

### BLEU Scores:-

| Translation Task | NLLB-200 | IndicTRANS | ChatGPT |
|---|---|---|---|
| **Eng to Hindi** | 0.6238 | 0.7011 | 0.6630 |
| **Hindi to Eng** | 0.6813 | 0.7599 | 0.6752 |
| **Guj to Hindi** | 0.6259 | 0.8062 | 0.2258 |
| **Hindi to Guj** | 0.5627 | 0.6463 | 0.2925 |

### ROUGE Scores for NLLB Translations:-

| Translation Task | Rouge-1 | Rouge-2 | Rouge-l |
|---|---|---|---|
| **Eng to Hindi** | {r: 0.5658, p: 0.6133, f: 0.5833} | {r: 0.3351, p: 0.3633, f: 0.3453} | {r: 0.5253, p: 0.5694, f: 0.5417} |

| | | | |
|---|---|---|---|
| **Hindi to Eng** | {r: 0.6384, p: 0.6611, f: 0.6294} | {r: 0.4009, p: 0.4145, f: 0.4041} | {r: 0.6076, p: 0.6283, f: 0.6135} |
| **Guj to Hindi** | {r: 0.5726, p: 0.6165, f: 0.5884} | {r: 0.3546, p: 0.3795, f: 0.3632} | {r: 0.5395, p: 0.5795, f: 0.5540} |
| **Hindi to Guj** | {r: 0.4843, p: 0.5456, f: 0.5066} | {r: 0.2558, p: 0.2899, f: 0.2678} | {r: 0.4633, p: 0.5213, f: 0.4843} |

## ROUGE Scores for IndicTRANS Translations:-

| **Translation Task** | **Rouge-1** | **Rouge-2** | **Rouge-l** |
|---|---|---|---|
| **Eng to Hindi** | {r: 0.6271, p: 0.6388, f: 0.6294} | {r: 0.3981, p: 0.4055, f: 0.3996} | {r: 0.5882, p: 0.5996, f: 0.5907} |
| **Hindi to Eng** | {r: 0.6994, p: 0.6967, f: 0.6948} | {r: 0.4786, p: 0.4752, f: 0.4741} | {r: 0.6698, p: 0.6670, f: 0.6653} |
| **Guj to Hindi** | {r: 0.7574, p: 0.7647, f: 0.7597} | {r: 0.5759, p: 0.5817, f: 0.5777} | {r: 0.7349, p: 0.7421, f: 0.7372} |
| **Hindi to Guj** | {r: 0.5423, p: 0.5477, f: 0.5417} | {r: 0.2938, p: 0.2943, f: 0.2923} | {r: 0.5189, p: 0.5240, f: 0.5183} |

## ROUGE Scores for chatGPT Translations:-

| Translation Task | Rouge-1 | Rouge-2 | Rouge-l |
|---|---|---|---|
| **Eng to Hindi** | {r: 0.5706, p: 0.5828, f: 0.5746} | {r: 0.3489, p: 0.3591, f: 0.3527} | {r: 0.5402, p: 0.5529, f: 0.5445} |
| **Hindi to Eng** | {r: 0.5929, p: 0.5889, f: 0.5883} | {r: 0.3667, p: 0.3635, f: 0.3633} | {r: 0.5630, p: 0.5591, f: 0.5586} |
| **Guj to Hindi** | {r: 0.0142, p: 0.0139, f: 0.0140} | {r: 0.0066, p: 0.0065, f: 0.0065} | {r: 0.0142, p: 0.0139, f: 0.0140} |
| **Hindi to Guj** | {r: 0.1979, p: 0.2134, f: 0.2016} | {r: 0.0783, p: 0.0830, f: 0.0799} | {r: 0.1897, p: 0.2037, f: 0.1928} |

From these Observations we can draw several conclusions about these models:-
(these conclusions that i have drawn are on the basis of these observations only which i got after running these models)

**NLLB-200 Model Strengths**:

In NLLB model i have observed that it demonstrates strong performance in translating between Indian languages, including English, Gujarati, and Hindi, as indicated by relatively high BLEU and ROUGE scores for this model.
Also it exhibits consistent recall, precision, and F-scores across different language pairs and ROUGE metrics, which shows us the robustness of this specific model.

**NLLB-200 Model Limitations**:-

Its overall performance is very good, but in some areas it can improve like the cases where more complex language pairs or translation are required.
maintaining higher precision and F-scores, especially for Handling nuances in translation, such as idiomatic expressions or cultural context, might pose challenges for the model.

**IndicTrans Model Strengths:**

IndicTrans demonstrates high-quality translation, especially for languages with limited resources, as indicated by its strong BLEU and ROUGE scores across various language pairs.
It excels in handling Indian languages and maintains good recall, precision, and F-scores, showcasing its effectiveness in translating between different language pairs.

**IndicTrans Model Limitations:**

While performing well, there may still be room for improvement, particularly in maintaining higher precision and F-scores consistently across all translation tasks.

Handling complex linguistic structures and contextual nuances might be challenging for this model, impacting translation quality in certain scenarios but its ot the case with every translation.

**ChatGPT Strengths:**

ChatGPT is not specifically designed for translation,but still it exhibits decent performance in translation tasks, especially considering its broader language capabilities.(for hindi to english and english to hindi) It shows reasonable recall and precision in some translation tasks, such as English to Hindi.

**ChatGPT Limitations:**

ChatGPT's performance in translation tasks appears to be less consistent compared to specialized translation models like IndicTrans, with significantly lower BLEU and ROUGE scores in certain cases. The model's lower recall, precision, and F-scores suggest potential challenges in accurately capturing and translating complex linguistic nuances and context.

Overall, these observations suggest that while all three models have their strengths and weaknesses, IndicTrans (NLLB) and IndicTrans appear to be better suited for translation tasks, especially across Indian languages, due to their specialized design and training on relevant corpora. ChatGPT, while capable to some extent, may not match the performance of dedicated translation models in certain scenarios.

**Comparison in between these 3 models translations and Learnings :-**

IndicTrans outperformed the distilled NLLB-200 model and ChatGPT in terms of BLEU and ROUGE scores for most language pairs involving Indian languages.
The performance of ChatGPT was comparable to IndicTrans for English-to-X and X-to-English translations, but it lagged behind for translations involving other Indian languages.
The distilled NLLB-200 model exhibited the lowest scores across all language pairs, likely due to its smaller size and the fact that it was not specifically optimized for Indian languages.
Model Evaluation:- We learned how to assess the performance of machine translation models using metrics like BLEU and ROUGE scores. This helped us understand the strengths and weaknesses of different models in translating between languages.
Dataset Handling:- Working with large datasets required careful organization and preprocessing. We gained practical experience in managing and preparing data for machine translation tasks, ensuring data quality and consistency.
Low-Resource Languages:- Translating low-resource languages presents unique challenges due to limited linguistic resources. We learned strategies for overcoming these challenges, such as data augmentation and transfer learning, to improve model performance.
Comparative Analysis:- Evaluating multiple models on the same datasets allowed us to compare their performance objectively. This taught us how to establish baselines, set performance benchmarks, and identify the most suitable model for specific translation tasks.

**NOTE:** I have selected Gujrati as the third language and chatgpt is not performing good on gujarati language and maybe chatgpt is more robustly trained for devanagari script languages thats why its not performing good for gujarati language(gujarati script do not comes under devanagari script)

Apart from chatgpt i have also tried gujarati language translations from CLAUD and it performed far better than CHATGPT

Overall, this assignment provided a valuable learning experience that enhanced our understanding of machine translation, dataset handling, and the challenges associated with working with low-resource languages.