Question 3:

1) NLLB Results

NLLB_Bleu_scores:

        english_to_marathi: **0.5335583164262447**,

        marathi_to_english: **0.594567484676005**,

        hindi_to_marathi: **0.551047276200799**,

        marathi_to_hindi: **0.572869802765165**

NLLB_rough_scores:

        english_to_marathi:

                rouge_1_score: **0.43169920183462335**,

                rouge_2_score: **0.18892563980629434**,

                rouge_L_score: **0.4075101502385872**

        marathi_to_english:

                rouge_1_score: **0.5676544691467368**,

                rouge_2_score: **0.3063370363329554**,

                rouge_L_score: **0.5312188059160582**

        hindi_to_marathi:

                rouge_1_score: **0.4495832064500809**,

                rouge_2_score: **0.20506214060500175**,

                rouge_L_score: **0.4205847370982619**

        marathi_to_hindi:

                rouge_1_score: **0.529671869879454**,

                rouge_2_score: **0.28502955383428435**,

                rouge_L_score: **0.48976134743956723**

2) IndicTrans Results:

ndicTrans_bleu_scores:

        english_to_marathi: **0.5769371555161382**,

        marathi_to_english: **0.6328086822526691**,

        hindi_to_marathi: **0.5647200934122635**,

        marathi_to_hindi: **0.576661062943658**

indicTrans_rough_scores:

        english_to_marathi:

                rouge_1_score: **0.47813130550373456**,

                rouge_2_score: **0.22648446830177682**,

                rouge_L_score: **0.45078793542969253**

        marathi_to_english:

                rouge_1_score: **0.6042580103571887**,

                rouge_2_score: **0.35540993002009713**,

                rouge_L_score: **0.5681003027304502**

        hindi_to_marathi:

                rouge_1_score: **0.4588458806370799**,

                rouge_2_score: **0.21372247161904093**,

rouge_L_score: **0.4329109563489027**

marathi_to_hindi:
     rouge_1_score: **0.533477388472777**,
     rouge_2_score: **0.2900066785699815**,
     rouge_L_score: **0.4931812951132034**

3) **Chatgpt Results**

chatgpt_bleu_scores:
     english_to_marathi: **0.40269466961916955**,
     marathi_to_english: **0.5740242540534669**,
     hindi_to_marathi: **0.41667575707339183**,
     marathi_to_hindi: **0.4779575870696223**

chatgpt_rough_scores:
     english_to_marathi:
          rouge_1_score: **0.2350136014863408**,
          rouge_2_score: **0.055611550796473466**,
          rouge_L_score: **0.21687840049860269**

     marathi_to_english:
          rouge_1_score: **0.5051004290518939**,
          rouge_2_score: **0.2614893352829392**,
          rouge_L_score: **0.463672019146262**

     hindi_to_marathi:
          rouge_1_score: **0.23415462661113395**,
          rouge_2_score: **0.044204121904987896**,
          rouge_L_score: **0.2090676085317885**

     marathi_to_hindi:
          rouge_1_score: **0.41375937362132065**,
          rouge_2_score: **0.18144257333783878**,
          rouge_L_score: **0.3739243498525802**

Observations and learnings:

As from above results, IndicTrans looks like the most reliable model, outperforming NLLB and Chatgpt in both Bleu and Rouge scores. NLLB is not as consistent as IndicTrans, but it is also performing reasonably well. But chatgpt lagging behind both these models, mainly in Bleu scores. We can say that the performance differences between these models can be due to various factors, like differences in model architecture, training data and preprocessing technique.

Also, authors of indicTrans model have created separate tokenizer. As we know that the way text is preprocessed and tokenized can affect translation quality. So, here IndicTrans could have more effective preprocessing techniques for indian languages. Hence it is giving better results than other models.