Question 3:

Comparison :

This report is showing comparison between translation of text files using NLLB Model, Indic-Trans Model and ChatGPT.Below given are the BLEU and ROUGE Score of each translation for each Model.

BLEU Score :

The BLEU (Bilingual Evaluation Understudy) score is a metric used to evaluate the quality of machine-translated text by comparing it to one or more human reference translations. It was proposed as a way to automate the evaluation process of machine translation systems.

BLEU measures the similarity between the candidate translation (the output of a machine translation system) and one or more reference translations (human-generated translations). It operates by comparing n-grams (sequences of n consecutive words) in the candidate translation to those in the reference translations.

ROUGE Score :

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is a set of metrics used to evaluate the quality of summaries or translations by comparing them to one or more reference summaries or translations. ROUGE focuses on measuring the overlap between the generated summary or translation and the reference summaries or translations.

NLLB Model :

Hindi to english:

Sentence BLEU Score: 0.6664176211854691

Rouge Score :

rouge-1 : {'r': 0.629112395790558, 'p': 0.6115796173794302, 'f': 0.6156453078616192}
rouge-2 : {'r': 0.38028558514177707, 'p': 0.3722125093260254, 'f': 0.3729146138365017}
rouge-l : {'r': 0.5966496128976457, 'p': 0.5808978454926305, 'f': 0.5843442732112495}

English to Hindi :

Sentence BLEU Score: 0.610053244943888

Rouge Score :

rouge-1 : {'r': 0.5920702184109471, 'p': 0.5558459075674409, 'f': 0.5682227027427238}
rouge-2 : {'r': 0.34583587610468425, 'p': 0.32485279527914884, 'f': 0.3317918936862638}
rouge-l : {'r': 0.5508588166470766, 'p': 0.5171752867492838, 'f': 0.5287685374339159}

Marathi to Hindi :

Sentence BLEU Score: 0.56066217477413

Rouge Score :

rouge-1 : {'r': 0.5382534786355637, 'p': 0.49651588224924587, 'f': 0.5109430120271112}
rouge-2 : {'r': 0.28188165547543786, 'p': 0.2607617194297417, 'f': 0.2678061750911307}
rouge-l : {'r': 0.49994147665166716, 'p': 0.46143442785486233, 'f': 0.47479278292823746}

Hindi to Marathi :

Sentence BLEU Score: 0.5461467974318344

Rouge Score :

rouge-1 : {'r': 0.4419535781555906, 'p': 0.41123943793303536, 'f': 0.42149200417237165}
rouge-2 : {'r': 0.19764071763290095, 'p': 0.1852154728263277, 'f': 0.18906147629141523}
rouge-l : {'r': 0.41128942761315024, 'p': 0.3837294746802353, 'f': 0.3928164624761462}

Indic-Trans Model:

Hindi to english :

BLEU Score: 0.7509905128899935

Rouge Score :

rouge-1 : {'r': 0.6792330438991676, 'p': 0.6837362307411765, 'f': 0.6782925579579643}
rouge-2 : {'r': 0.4475947157750621, 'p': 0.4531815130513081, 'f': 0.44787541854447943}
rouge-l : {'r': 0.6463267876918584, 'p': 0.6508646447333915, 'f': 0.6455910560254102}

English to hindi :

BLEU Score: 0.6948488300274981

Rouge Score :

rouge-1 : {'r': 0.6312974415552479, 'p': 0.6236171066349142, 'f': 0.6240577378456516}
rouge-2 : {'r': 0.3986220928405689, 'p': 0.3949067197710193, 'f': 0.39462848175821374}
rouge-l : {'r': 0.5960898710186854, 'p': 0.5884980608189196, 'f': 0.5891420496399068}

Marathi to hindi :

BLEU Score: 0.6117528744937303

Rouge Score :

rouge-1 : {'r': 0.5411127663945184, 'p': 0.5282191814299513, 'f': 0.5299542390222854}
rouge-2 : {'r': 0.2899827287436648, 'p': 0.2836610914281554, 'f': 0.2842049112678615}
rouge-l : {'r': 0.5008516425345846, 'p': 0.48826049614286055, 'f': 0.49025361044416915}

Hindi to marathi :

BLEU Score: 0.6102149802393331

Rouge Score :

rouge-1 : {'r': 0.4643810070623246, 'p': 0.46376841123297396, 'f': 0.46043185167056927}
rouge-2 : {'r': 0.21426397146689083, 'p': 0.21625876776118724, 'f': 0.21353291112366585}
rouge-l : {'r': 0.43821989754506757, 'p': 0.43817663692225806, 'f': 0.43475286577210537}

ChatGPT :

Hindi to english :
BLEU Score: 0.7010350641310221
Rouge Score :
rouge-1 : {'r': 0.6540950310439474, 'p': 0.6424568561813144, 'f': 0.6443898180930964}
rouge-2 : {'r': 0.4065279020944656, 'p': 0.3978530591841267, 'f': 0.39929385766846354}
rouge-l : {'r': 0.6163150052561815, 'p': 0.6068014478641414, 'f': 0.6079362112386781}

English to hindi :
BLEU Score: 0.6087497144144103
Rouge Score :
rouge-1 : {'r': 0.5457438548379725, 'p': 0.5225178257930578, 'f': 0.5318037220198154}
rouge-2 : {'r': 0.29074108795911857, 'p': 0.277388777491335, 'f': 0.2824066105138424}
rouge-l : {'r': 0.5184900847518494, 'p': 0.49682734827951536, 'f': 0.5055046775268505}

Hindi to marathi :
BLEU Score: 0.41635099662376235
Rouge Score :
rouge-1 : {'r': 0.2587855839438825, 'p': 0.2456614116018141, 'f': 0.2490714803991816}
rouge-2 : {'r': 0.0640758936289627, 'p': 0.06652414421278519, 'f': 0.0645416797374231}
rouge-l : {'r': 0.24223161163031948, 'p': 0.2318914430083161, 'f': 0.23432072995617861}

Marathi to hindi :
BLEU Score: 0.5250565219714854
Rouge Score :
rouge-1 : {'r': 0.4320744258886153, 'p': 0.41752748746071533, 'f': 0.4201514222064645}
rouge-2 : {'r': 0.2065873548773089, 'p': 0.19672873613675873, 'f': 0.20042915620464058}
rouge-l : {'r': 0.40107638772635673, 'p': 0.3895303078257649, 'f': 0.39100823004225743}

Overall Observations:

The Indic-Trans model generally outperforms both the NLLB Model and ChatGPT across all translation tasks, with consistently higher BLEU and Rouge scores.
ChatGPT performs reasonably well, especially in Hindi to English translation, but lags behind the Indic-Trans model.
The NLLB Model shows competitive performance but falls slightly behind the Indic-Trans model in most cases.
Marathi to Hindi translation appears to be a challenging task for all models, with lower scores across the board.

Conclusion:

The comparison of model scores reveals distinct performance discrepancies across various translation tasks. While the NLLB Model demonstrates competitive performance in Hindi to English translation, its efficacy diminishes in other tasks, notably Marathi to Hindi. In contrast, the Indic-Trans model consistently outperforms both the NLLB Model and ChatGPT, exhibiting robustness across all translation tasks. ChatGPT, while showing promise, tends to lag behind the Indic-Trans model, particularly struggling with Marathi translations. These findings underscore the importance of continual refinement and development of translation models, especially in handling diverse language pairs like Hindi, English, and Marathi. Further research efforts are warranted to enhance model capabilities, particularly in addressing challenges related to low-resource languages and achieving broader generalization. By leveraging a combination of evaluation metrics like BLEU and Rouge scores, researchers can gain comprehensive insights into model performance, ultimately driving advancements in machine translation technology.