

General Observations

The Facebook NLLB-200 model showcased variable translation accuracy, with some instances of high precision but generally low BLEU scores, suggesting it may struggle with complex or domain-specific texts.

The Helsinki-NLP models demonstrated consistent yet limited translation capability, with slightly better performance on standardized datasets compared to custom texts.

Google Translate and Yandex Translate mirrored the Helsinki-NLP models' performance on specialized texts, suggesting possible difficulties with complex or nuanced content.

During the testing phase model best behaved on GPU computing powers also batching of the sentences show slight improvement on the speed of the model and did not resulted in different BLUE Scores. However difference in speed was not significant around 10% from 1 sentence batch to 16 sentence batch. Also batches have to take in account that models have limited token size around 512 which is not enough for the large text with contextual hard translation. The consistently low BLEU scores for custom texts across all models and services highlight the inherent challenges in machine translation, especially with sophisticated or less-common language pairs.

Analysis of Facebook NLLB-200

Translation from Russian to Various Languages:

Exhibited inconsistent outcomes with certain translations achieving perfect BLEU scores of 100, indicating potential disparities in translation quality.

Translation from English to Various Languages:

A notably low average BLEU score of 0.77 suggests significant challenges in translations from English, perhaps due to misalignments with the model's target languages.

Russian to English Translations:

Uniform BLEU scores of 1.11 across various batch sizes hint at a consistent but limited efficacy in translating from Russian to English.

English to Russian Translations:

Consistent BLEU scores of 1.17 indicate limited capability in translating from English to Russian.

Custom News and Fiction Text Translations:

BLEU scores below 0.2 for translations from both English to Russian and Russian to English on custom texts point to difficulties in handling complex or specialized content.

Insights from Helsinki-NLP Models

Custom Text Translations:

Translations from English to Russian and Russian to English show limited proficiency with BLEU scores of 0.11 and 0.28, respectively.

Translations using the FLORES Dataset:

Demonstrated improved performance with BLEU scores of 5.8 for translations from both Russian to English and English to Russian.

Evaluations of Google Translate and Yandex Translate

Custom English to Russian Text Translations:

Google Translate and Yandex Translate both yielded low BLEU scores of 0.12 and 0.11, indicating challenges in accurately translating nuanced content.