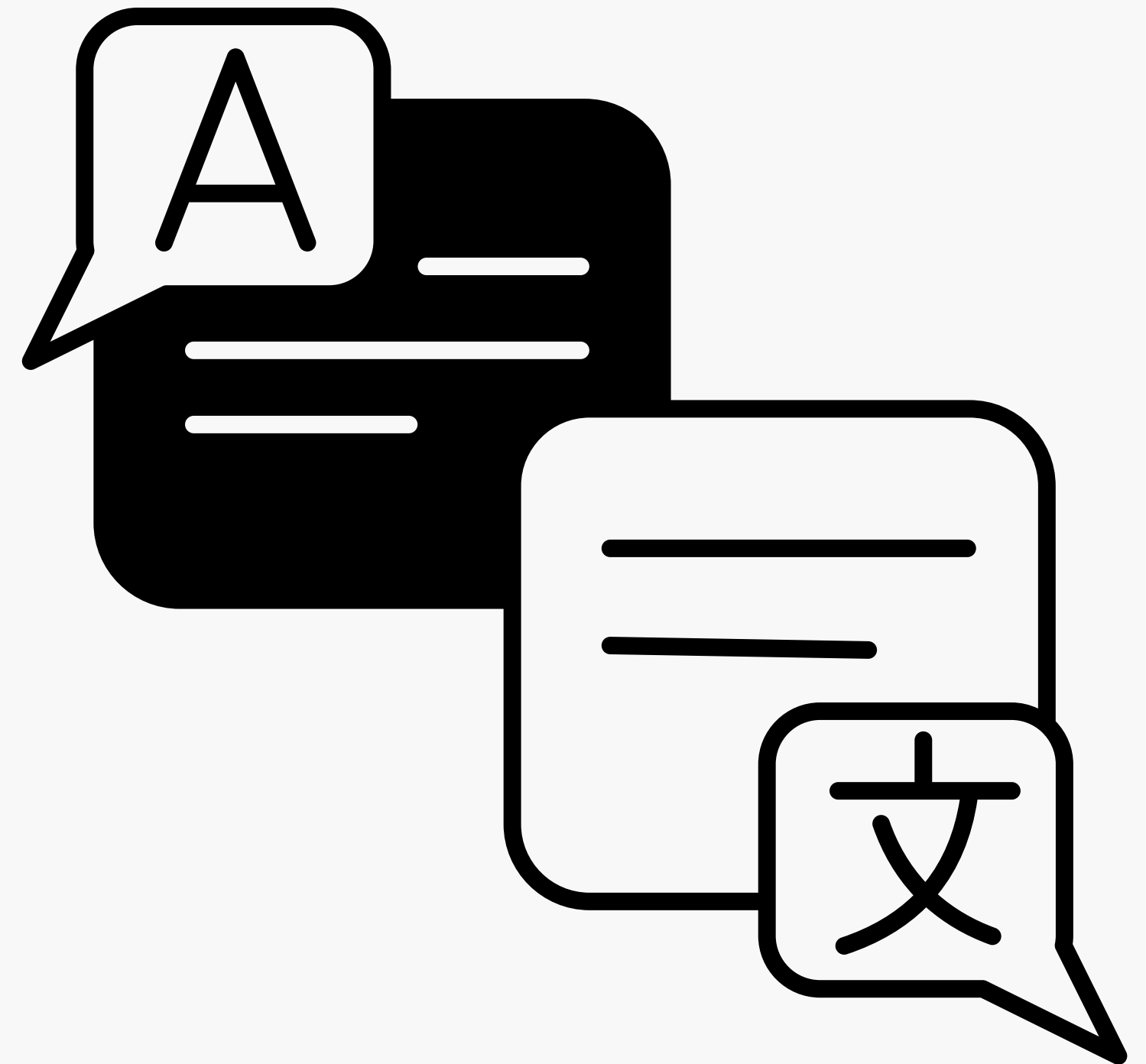


Machine Translation

English - Punjabi

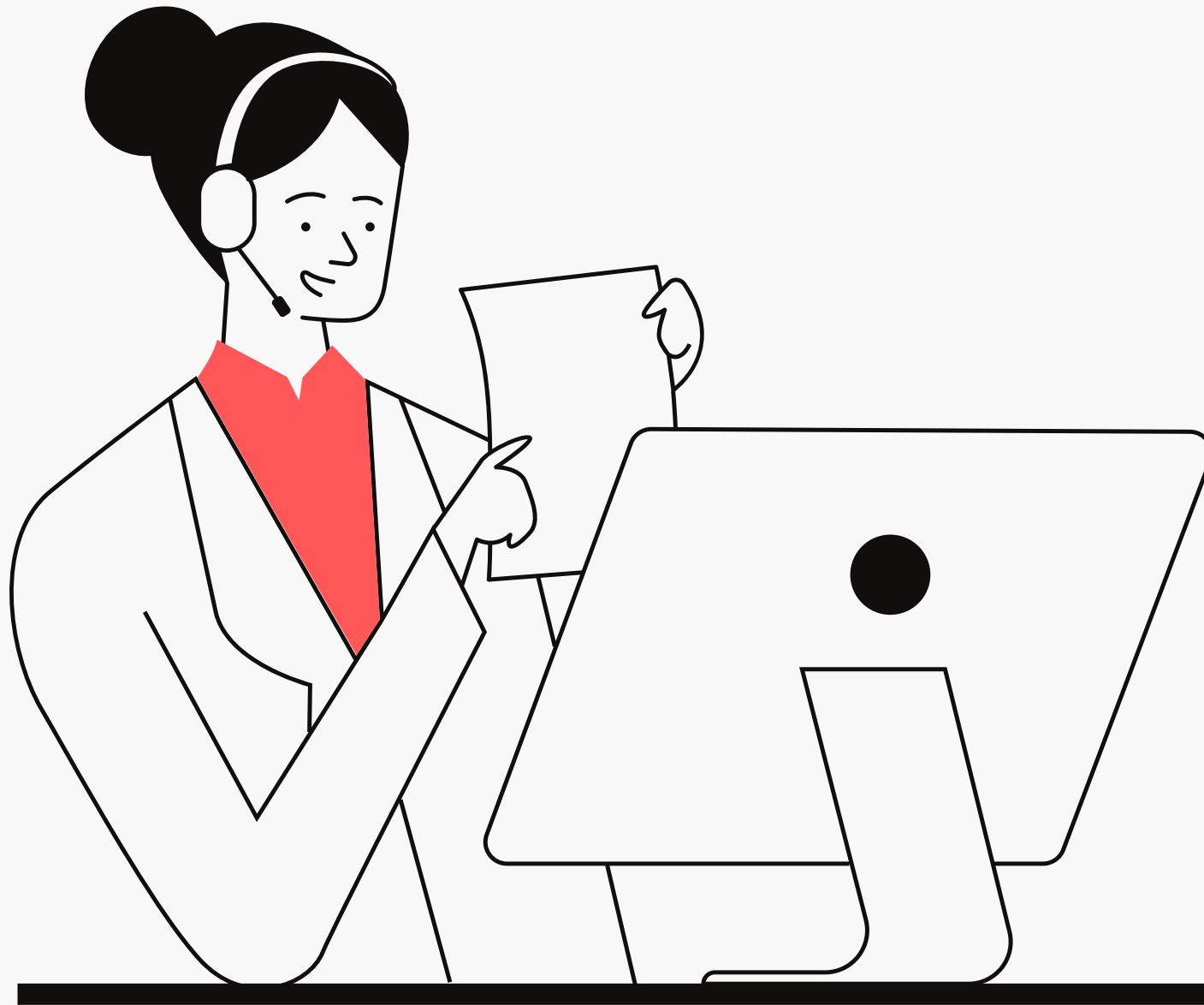
Team - Redlock

Aung 6411325
Tanat 6410381

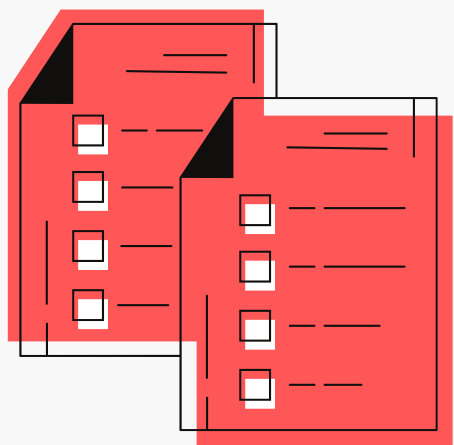


What is Machine Translation?

Machine translation is the automated process of converting text or speech from one language into another using computer algorithms and models to facilitate cross-lingual communication and understanding.



Challenges in Machine Translation from English to Punjabi



Structural and morphological divergence

English and Punjabi have different grammatical structures and word orders.

Examples:

- I love dog. - ਮੈਨੂੰ(I) ਕੁੱਤਾ(dog) ਪਸੰਦ(love) ਹੈ।
- I love dogs. - ਮੈਨੂੰ(I) ਕੁੱਤੇ(dogs) ਪਸੰਦ(love) ਹਨ।

Politeness and Formality

Punjabi has different levels of formality and politeness that are expressed through verb conjugations and vocabulary choices.

Examples:

- You - ਤੁਸੀਂ (tusī) Formal Pronouns
ਤੂੰ (tū) Informal Pronouns
- Do - ਕਰੋ (karo) Formal
ਕਰ (kar) Informal

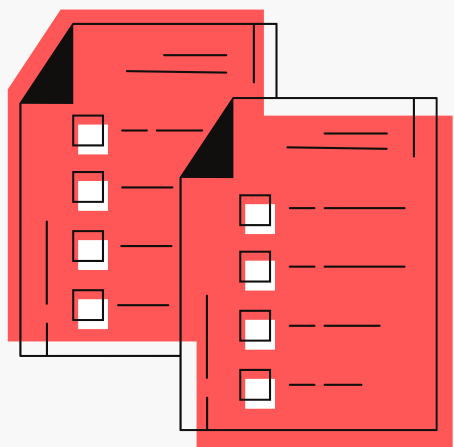
Challenges in Machine Translation from English to Punjabi

Idiomatic Expressions

Both languages have their own idiomatic expressions and cultural nuances, which may not have direct equivalents in the target language.

Low resource language

Limited amount of data and resources available for developing machine translation systems for this language pair.



Existing Techniques Used for MT Task

- **Neural Machine Translation**

NMT uses deep learning models, such as recurrent neural networks (RNNs) and transformers, to learn translation patterns from data.

- **Statistical Machine Translation**

SMT uses statistical models to learn translation patterns from large parallel corpora of English and Punjabi text. These models estimate the probability of a translation given a source sentence.

- **Convolutional sequence to sequence architecture**

The architecture comprises a Gated Linear Unit (GLU) and a Multi-Hop attention mechanism. GLU manages information flow for improved translations, while Multi-Hop attention enables repeated sentence analysis.



- **Bilingual Evaluation Understudy**

BLEU measures the overlap of n-grams (word sequences) between the machine-generated translation and one or more reference translations.

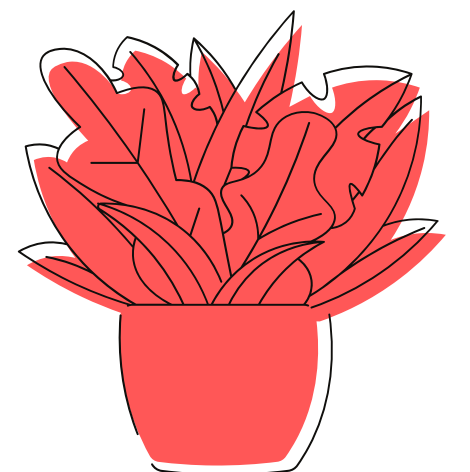
- **Metric for Evaluation of Translation with Explicit Ordering**

METEOR is another metric that considers precision, recall, stemming, synonymy, and word order. It tries to address some of BLEU's limitations.

- **Human Assessment**

Human evaluators rate translations for fluency, adequacy, and overall quality

How MT Models are Evaluated

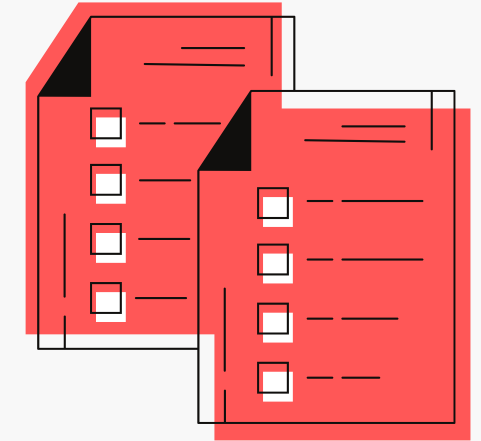


Is Automated Evaluation Credible?

Automated evaluation metrics like BLEU, METEOR, etc., provide a quick and objective way to compare different MT systems and track their progress over time. They are valuable for researchers and developers during model development and optimization. However, they have limitations, such as insensitivity to fluency and meaning, and they may not align perfectly with human judgment.

Automated metrics are credible for relative comparisons and tracking improvements but should not be the sole measure of translation quality.

Citations



- Patel, R. N., Pimpale, P. B., & Sasikumar, M. (2019). Machine translation in Indian Languages: Challenges and resolution. *Journal of Intelligent Systems*, 28(3), 437–445.
<https://doi.org/10.1515/jisys-2018-0014>
- Bansal, M., & Lobiyal, D. K. (2021). Multilingual sequence to sequence convolutional machine translation. *Multimedia Tools and Applications*, 80(25), 33701–33726.
<https://doi.org/10.1007/s11042-021-11345-6>
- Premjith, B., Kumar, M. A., & Soman, K. P. (2019). Neural Machine Translation System for English to Indian language translation using MTIL parallel corpus. *Journal of Intelligent Systems*, 28(3), 387–398. <https://doi.org/10.1515/jisys-2019-2510>