- Dataset:
    - OpenSLR
    - fleur
    - huggingface => ...
- Voice-Voice
    - Approach1:
        - Voice - Voice => high error, data
        - Stacked:
            - Voice to text
            - translate
            - TTS
- STT TTS:
    - whisper
    - mms
    - Nepali:
        - CNN-Transformer <= Shishir poudel
- Translation:
    - LSTM : Assignment <=

Presentation: - Introduce - Approach - Dataset

```
- STT, TTS -- CNN-Transformer
    - whisper, mms => huggingface ---

- Translation:
    - LSTM <= RNN

- Backtranslation <= Synthetic data
```

- English-Nepali
    - Nep => translate to english (quality..)
    - Art-Eng ~ Good nepali ---
- X-transformer => no. of encoder decoder change ---

=> BLEU score, CER (Character Error Rate), MOS

- Own voice
- Youtube video => https://www.youtube.com/watch?v=CZPmbAs5U24&t=38s
- Song => https://www.youtube.com/watch?v=DTlmuPNXGW0