

चित्रण: An Automated Festive Poster Generator with Wishes in Nepali Language

Team members:

Kristina Bhandari (THA077BCT022)

Kristina Ghimire (THA077BCT023)

Pradeepti Dongol (THA077BCT033)

Namita Bhatta (THA077BCT048)

Supervised By:

Er. Saroj Shakya

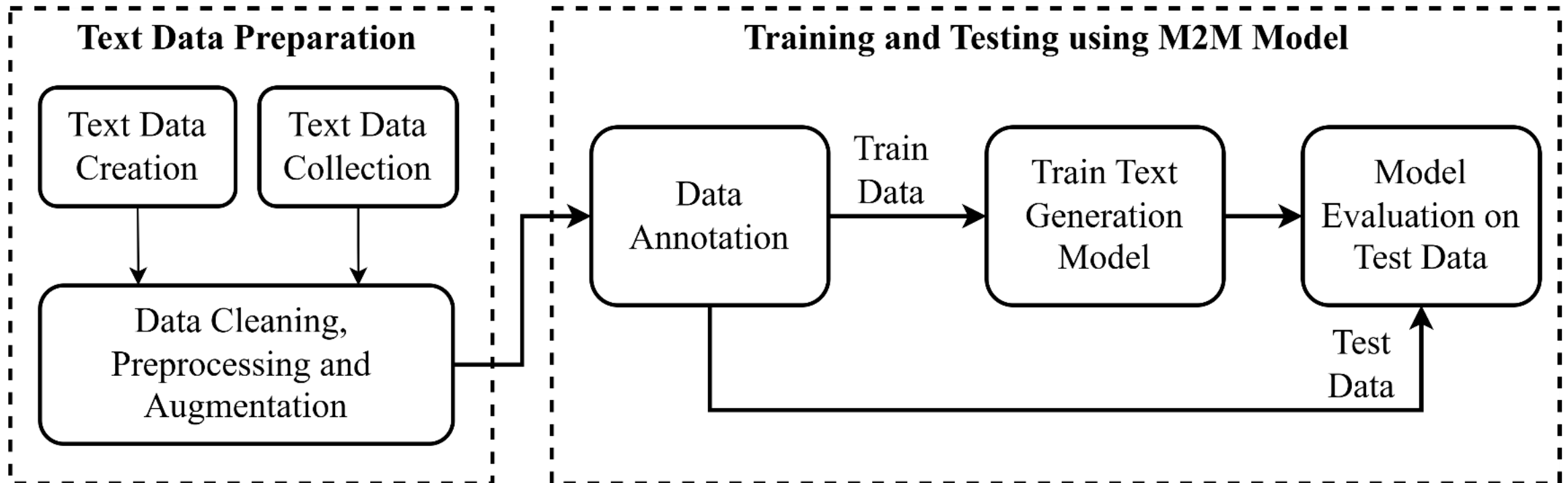
Department of Electronics and Computer Engineering
Institute of Engineering, Thapathali Campus

July, 2024

Objectives

- To analyze input prompt to extract festival themes, then generate concise Nepali short title.
- To generate a Nepali festival-themed image and integrate it with a styled title in Nepali to create a digital poster.

Methodology (System Block Diagram)



Methodology [Cont.]

(Text Dataset Preparation)

- Collected manually from different sources such as social media posters, greeting cards, and so on.
- Dataset contains text prompt in English and its corresponding wishes in Nepali font.
- So far, we have collected a dataset comprising 1095 pairs of prompts and wishes, where each prompt and wish is limited to 25 tokens in length.

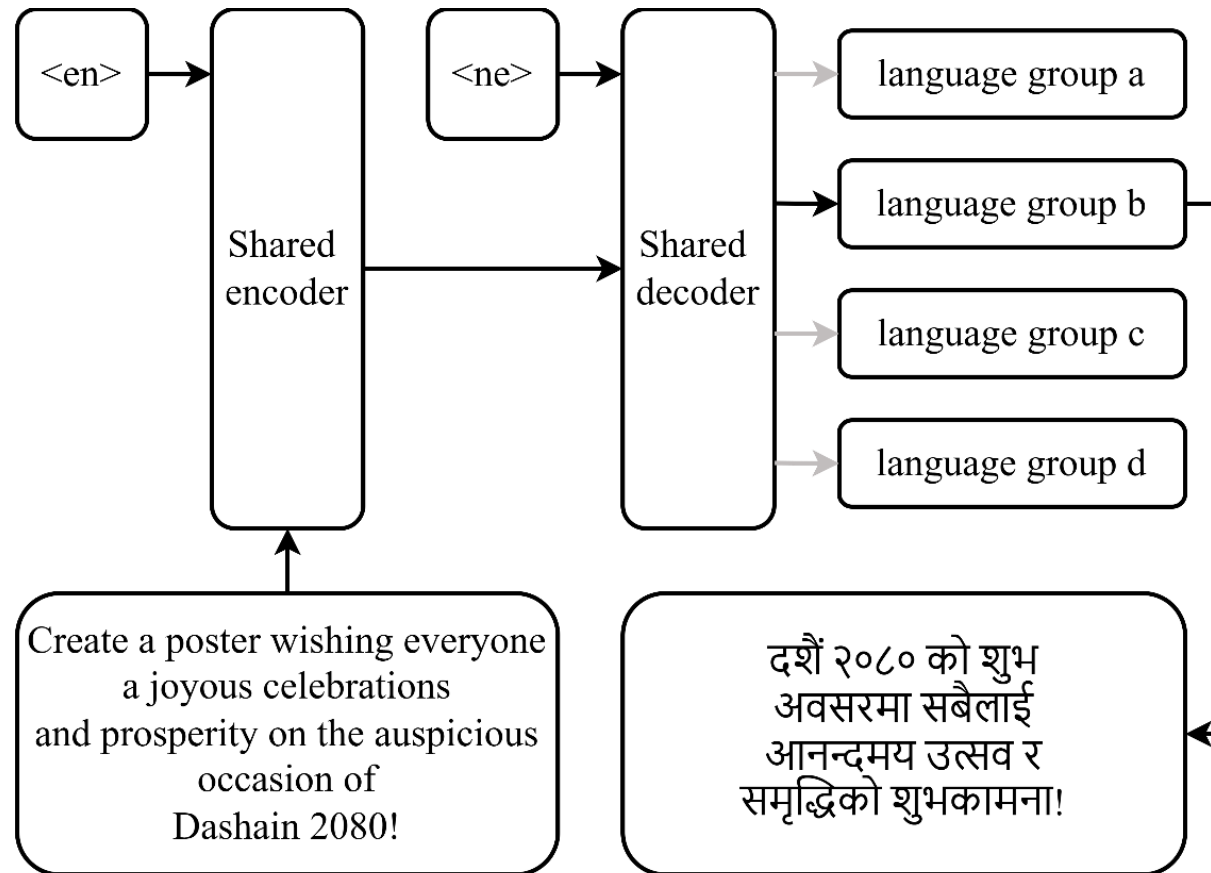
Methodology [Cont.]

(M2M-100 Model)

- Transformer-based model that can directly translate between 100 languages
- Uses 12 encoder layers and 12 decoder layers with 1024 hidden dimensions and 16 attention heads
- Pre-trained on a large-scale multilingual corpus
- Outperforms bilingual models and English-centric models

Methodology [Cont.]

(M2M-100 Model)

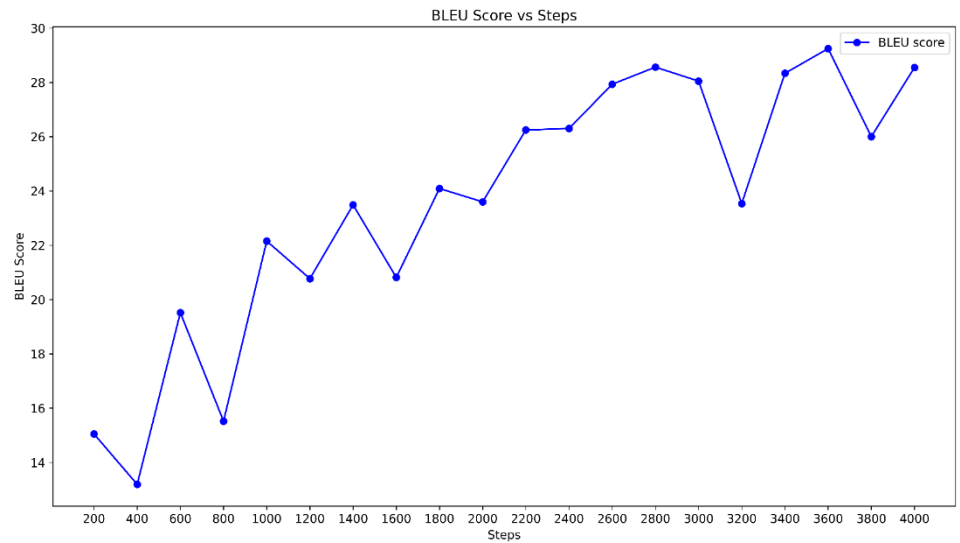
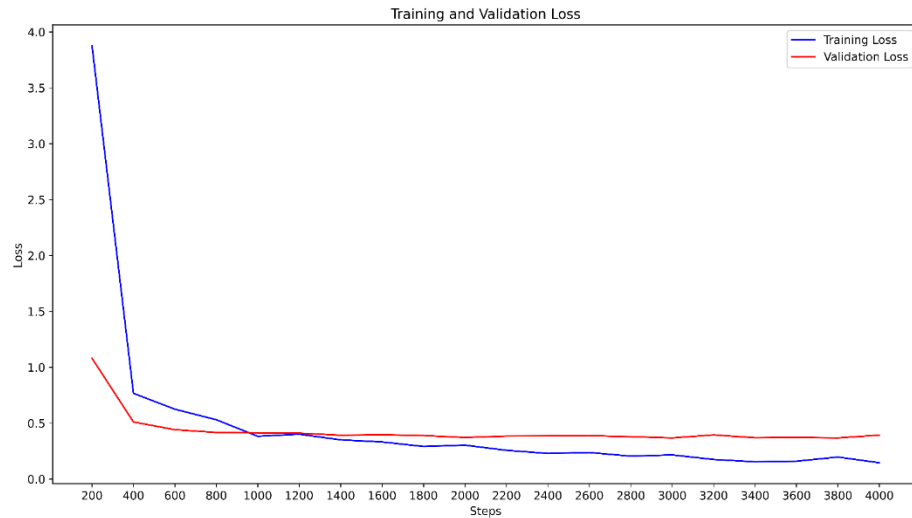


Methodology [Cont.]

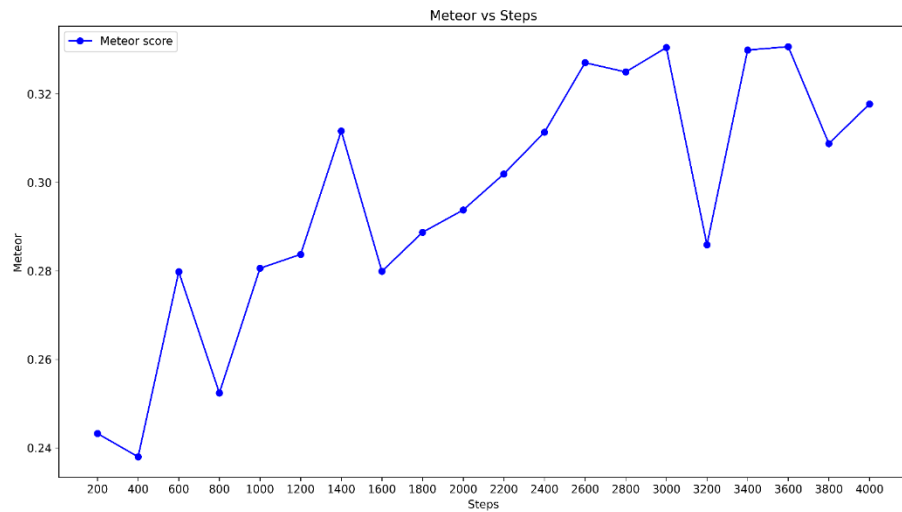
(Title Generation)

- For title generation, a pre-trained M2M-100 model is fine-tuned on our dataset.
- BLEU score, METEOR score, and TER score are used for evaluation.
- The M2M-100 model takes an English prompt and generates festival wishes in the Nepali language.

Results

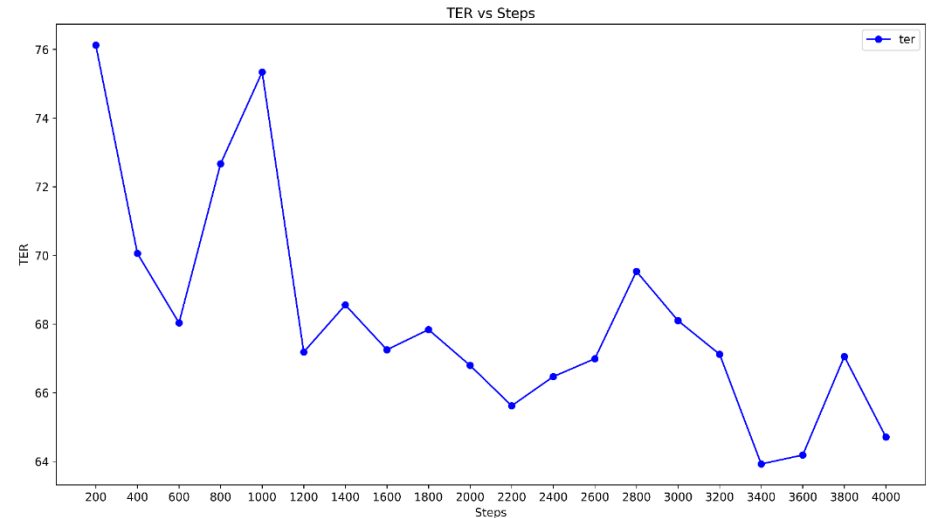


Loss Plot



Meteor Score Plot

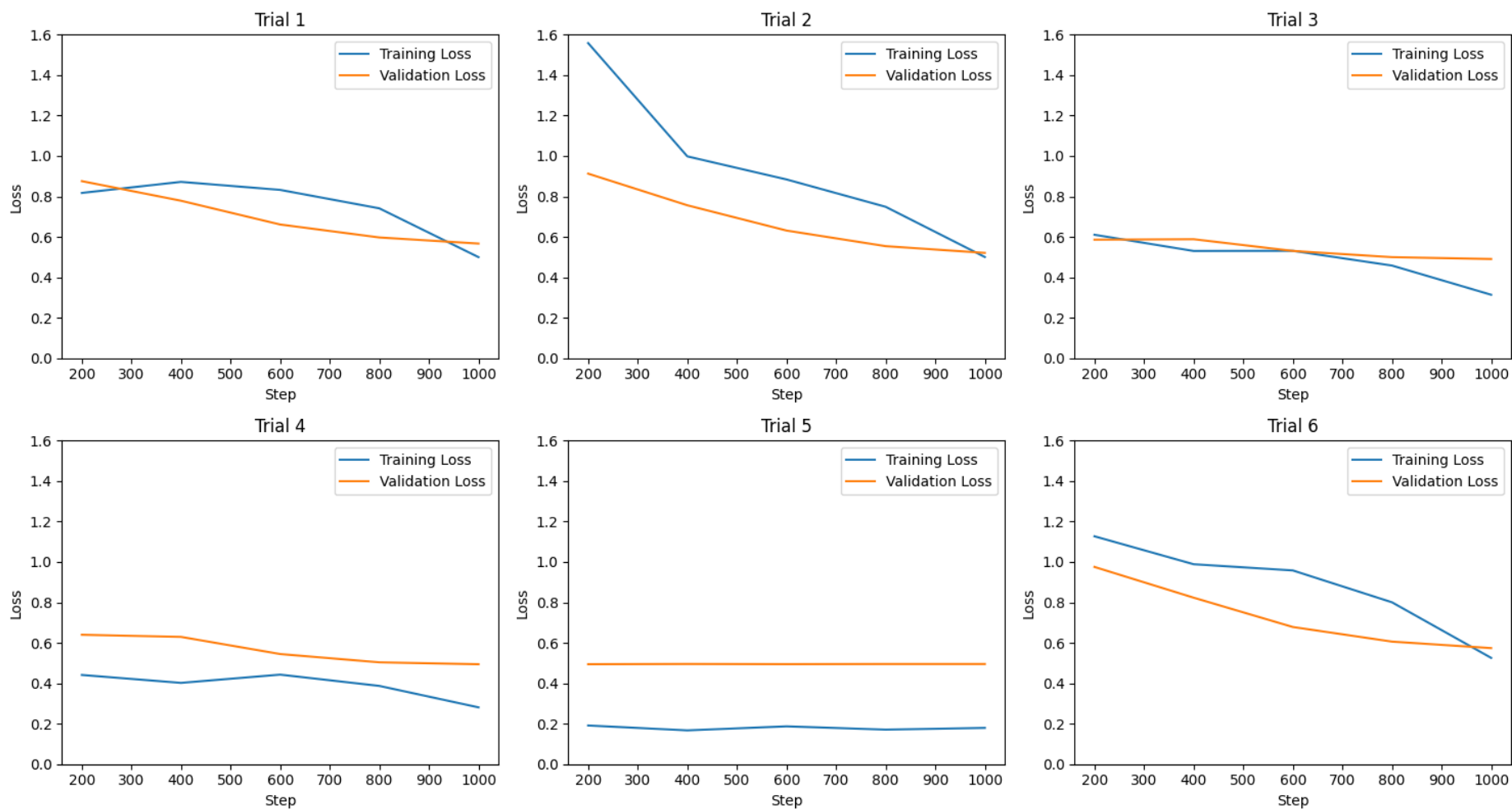
BLEU Score Plot



TER Score Plot

Results [Cont.]

(Parameter value selection using Optuna)



Results [Cont.]

(Text Generation)

Input Text: Create a Holi poster 2070

Output: Theme extracted

Output: Year corrected

Output: Year extracted

Output: ['सम्पूर्ण नेपालीलाई होली २०७० को पावन अवसरमा हार्दिक शुभकामना। यो पर्वले सबैको जीवनमा नयाँ उमंग र उत्साह ल्याओस्।']

Input Text: Wish healthy and happy dashain 2070 to family and friends

Output: Theme extracted

Output: Year corrected

Output: Year extracted

Output: ['परिवार र साथिहरूलाई २०७० को दशैंमा स्वास्थ्य र खुशीको शुभकामना!']

Results Analysis

- On training 1095 text dataset using M2M model, we achieved the maximum of 28.5 BLEU score.
- Training loss decreases faster but evaluation loss almost remains constant after certain steps.
- Obtained an maximum of 33% of Meteor score.
- TER decreased from 76.12 to 63.92.

Results Analysis [Cont.]

- The parameters that minimize evaluation loss were preferred.
- Use of Optuna suggested the following as best parameters to prevent overfitting:
 - Learning Rate: 0.0001753635227822039
 - Weight Decay: 0.19206671304964318

Remaining Tasks

- Increasing the text dataset
- Finetuning M2M model to obtain better scores.
- Image data collection
- Finetuning LDM
- Creating a user-friendly interface

References

- [1] A. Fan et al., “Beyond english-centric multilingual machine translation,” *Journal of Machine Learning Research*, vol. 22, no. 107, pp. 1–48, 2021.
- [2] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10684–10695.

References

- [1] A. Fan et al., “Beyond english-centric multilingual machine translation,” *Journal of Machine Learning Research*, vol. 22, no. 107, pp. 1–48, 2021.
- [2] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10684–10695.