MIT | Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

English Grammar Correction

Project Guide:

Dr. Diptee Ghusse

Dr. Sunita Barve

Group Members:

Rohan Agrawal (202201040208)

Rohit Jagtap (202201040048)

Rohan Wagh(202201040090)

Parth Kulkarni (202201040007)

INDEX

RESEARCH PAPER SUMMARY

1. **AIM :** To build an accurate grammar correction system using deep learning.

2. **OBJECTIVES:**
- Identify limits of traditional grammar tools.
- Develop a deep learning-based correction model.
- Use seq2seq with attention for better context.
- Evaluate model performance on benchmark datasets.

3. **PROBLEM STATEMENT :** Traditional grammar tools rely on fixed rules and lack context understanding, leading to low accuracy. A data-driven deep learning approach is needed for better correction.

Methodology

1.Model Architecture:

-> Utilized a Transformer-based encoder-decoder structure.

2.Tokenization & Preprocessing:

-> Applied WordPiece tokenization for input and output sequences.

3.Input Representation:

-> Input sentences passed to BERT to generate contextualized token embeddings.

-> Decoder receives shifted target sentences for training.

Methodology

4.Training Strategy:

-> Used teacher forcing to guide decoder training.

-> Employed cross-entropy loss as the optimization objective.

5.Data Used:

-> Trained and evaluated on public grammar correction datasets (e.g., CoNLL-2014, JFLEG).

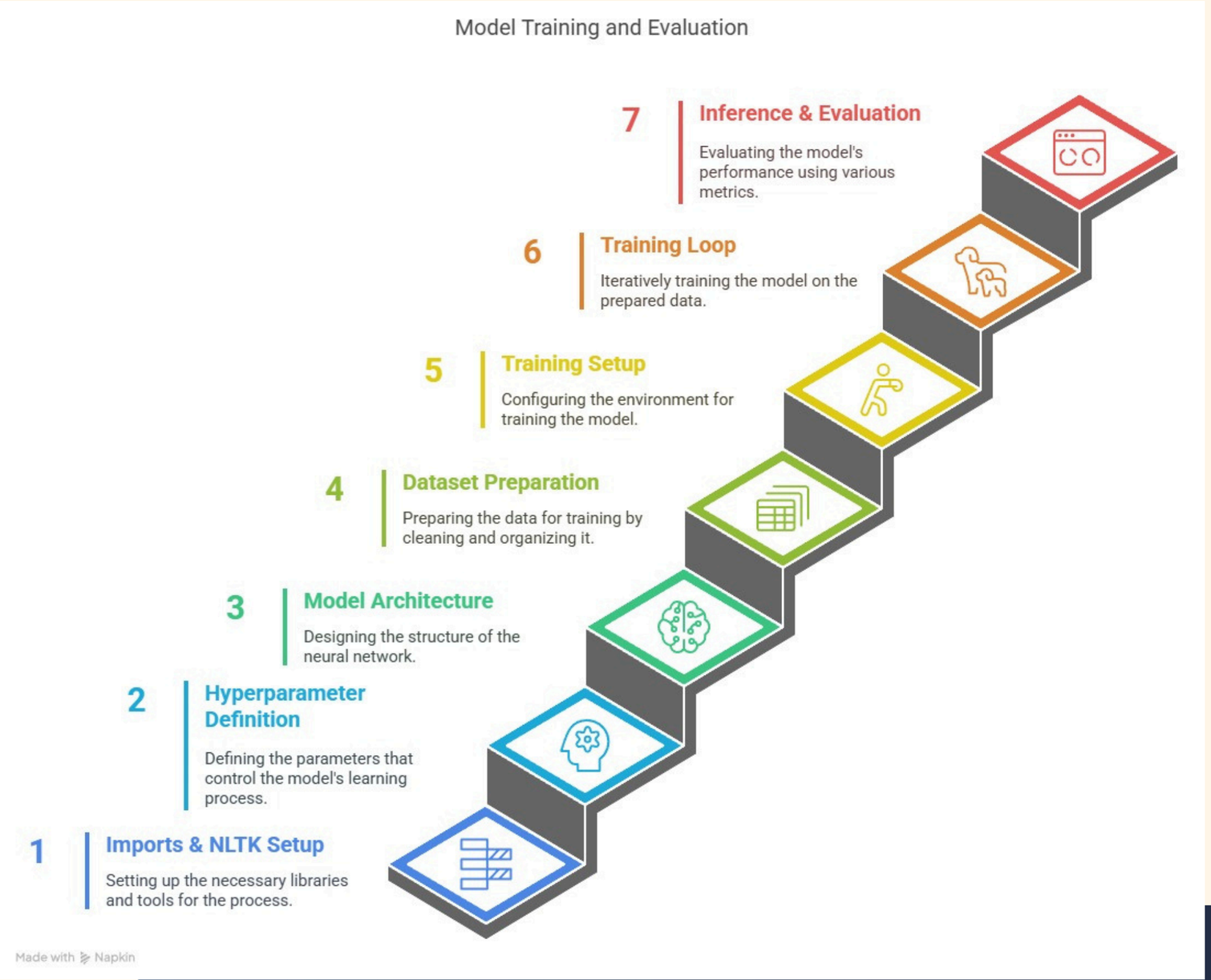
6.Evaluation Metrics:

-> Evaluated using Precision, Recall, F0.5-score, and GLEU score for effectiveness.

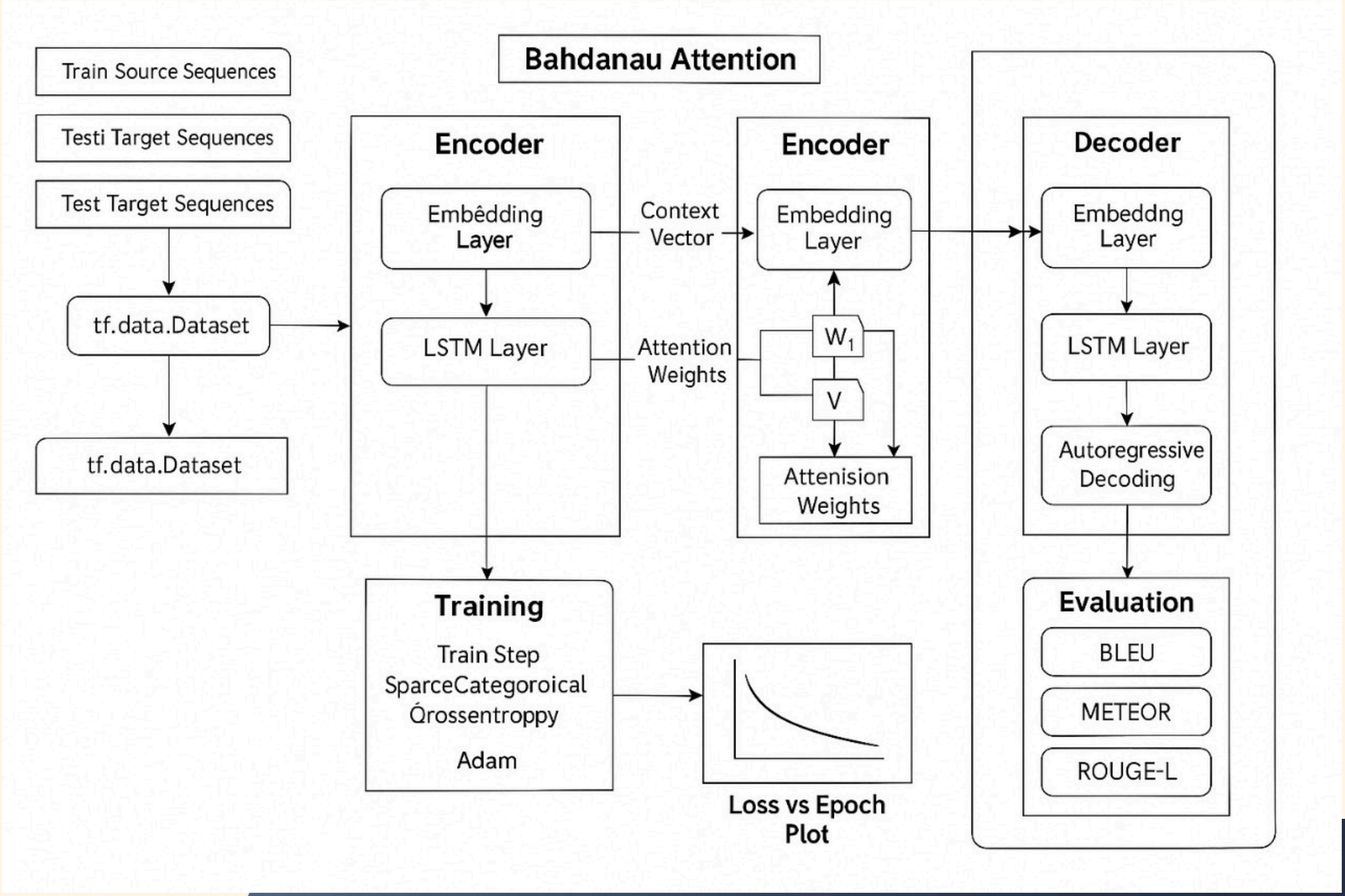
7.Inference:

-> Applied beam search during inference to generate corrected sentences

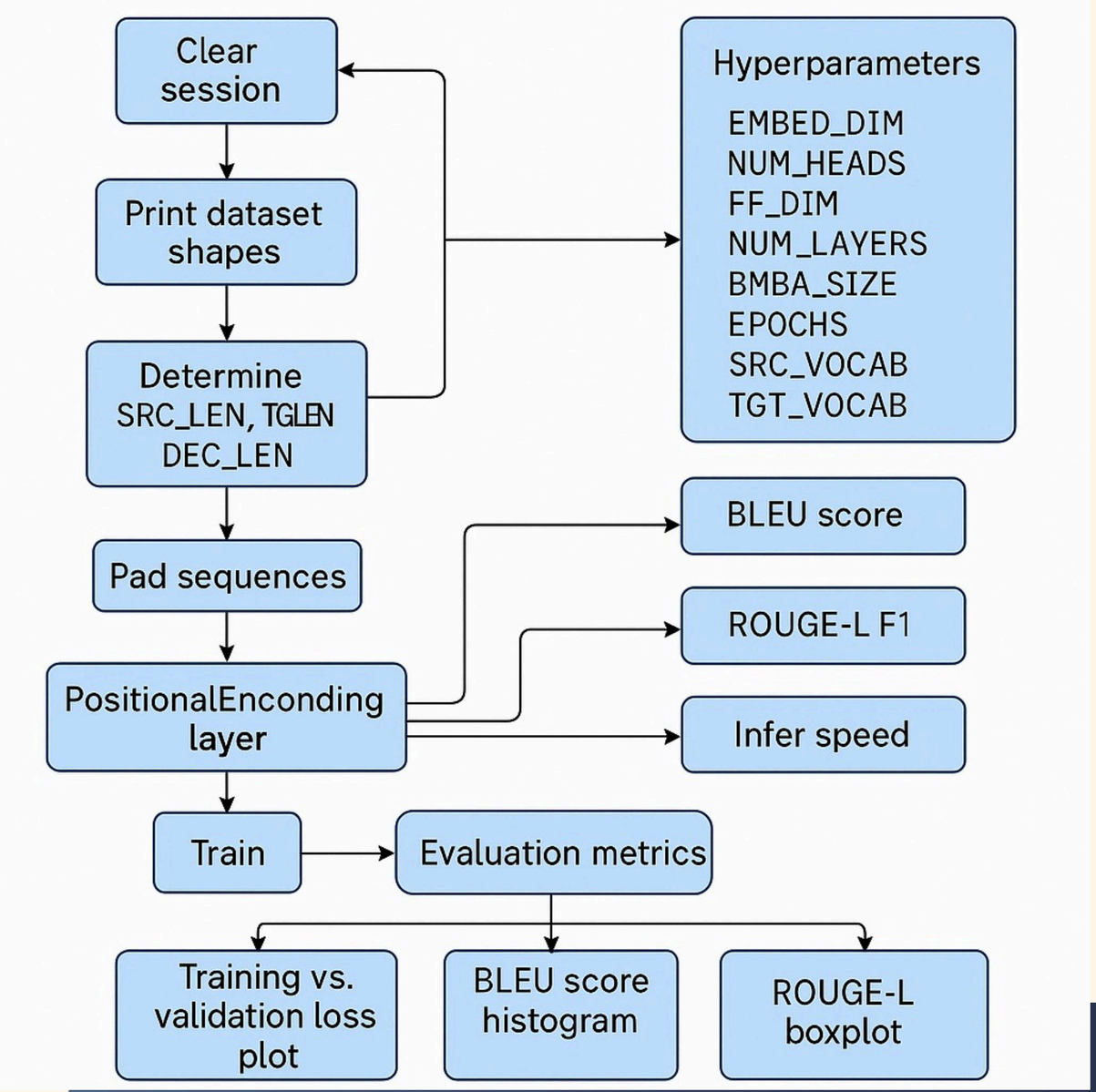
MODEL ARCHITECTURE (WITHOUT ATTENTION)



MODEL ARCHITECTURE (WITH ATTENTION)



MODEL ARCHITECTURE (SELF ATTENTION)



DATASET DESCRIPTION

Dataset Name: JFLEG Test Set

- Format: CSV (Comma-Separated Values)
- Purpose: Created for training and evaluating grammar correction models in Natural Language Processing (NLP).
- Total Records: 10,000
- Columns:
- Error Type – Category of the grammatical mistake.
- Ungrammatical Statement – The incorrect sentence containing grammar issues.
- Standard English – The corrected version of the sentence.

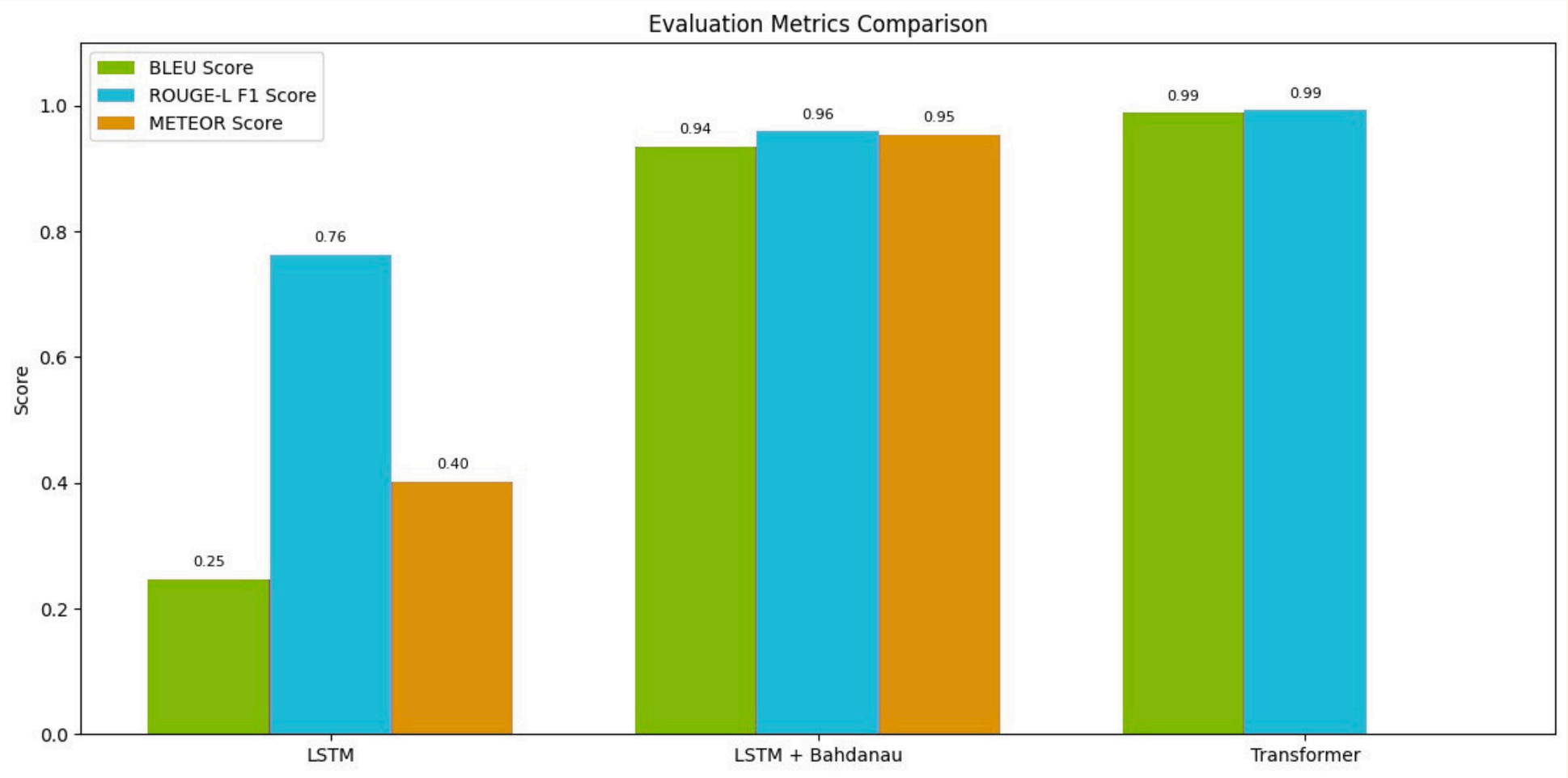
Use Cases:

- Grammar correction models (NLP)
- Grammar error detection and correction tools
- Language learning applications

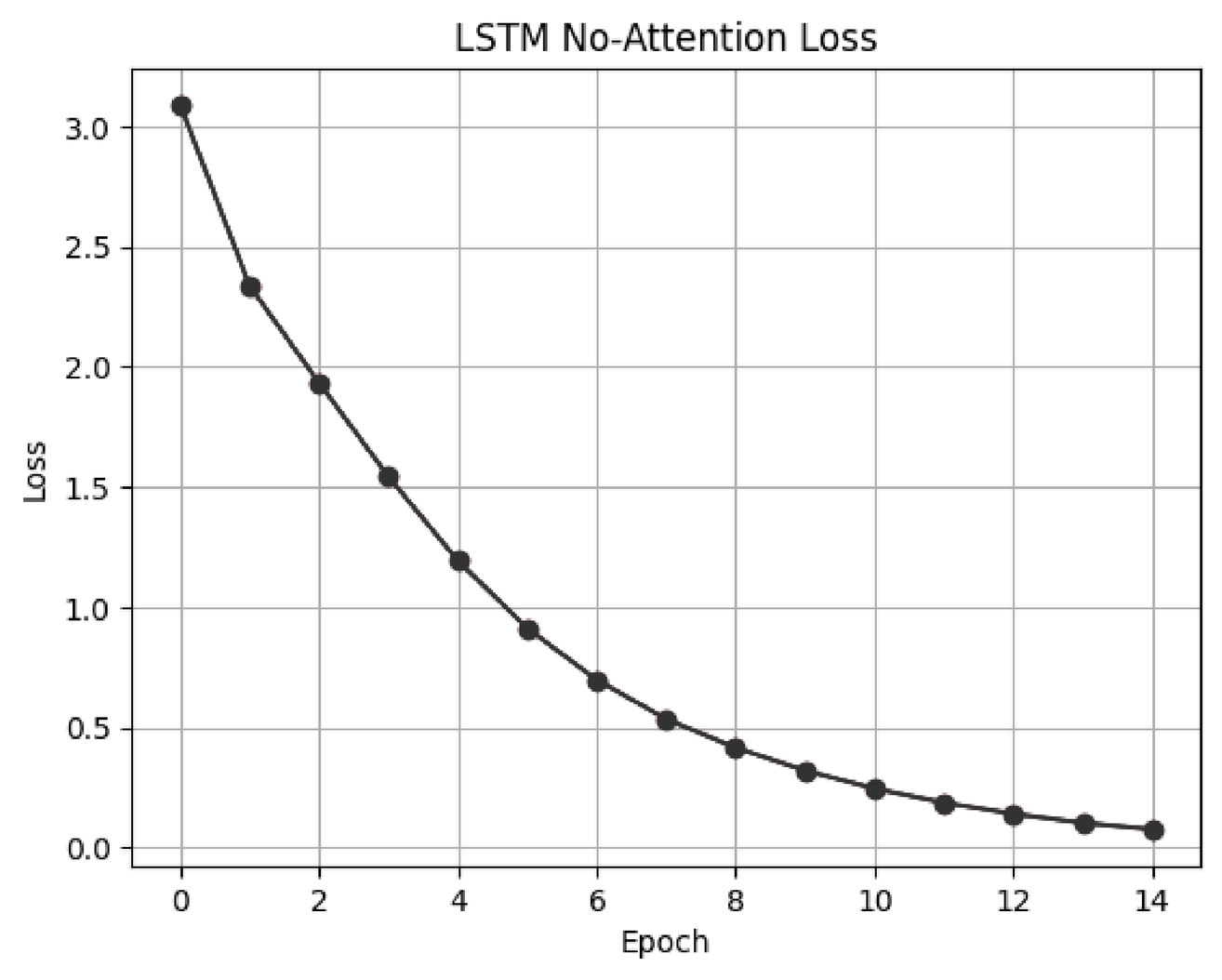
Ideal For:

- Training supervised machine learning models for sentence correction.
- Fine-tuning transformer-based models like BERT or T5.

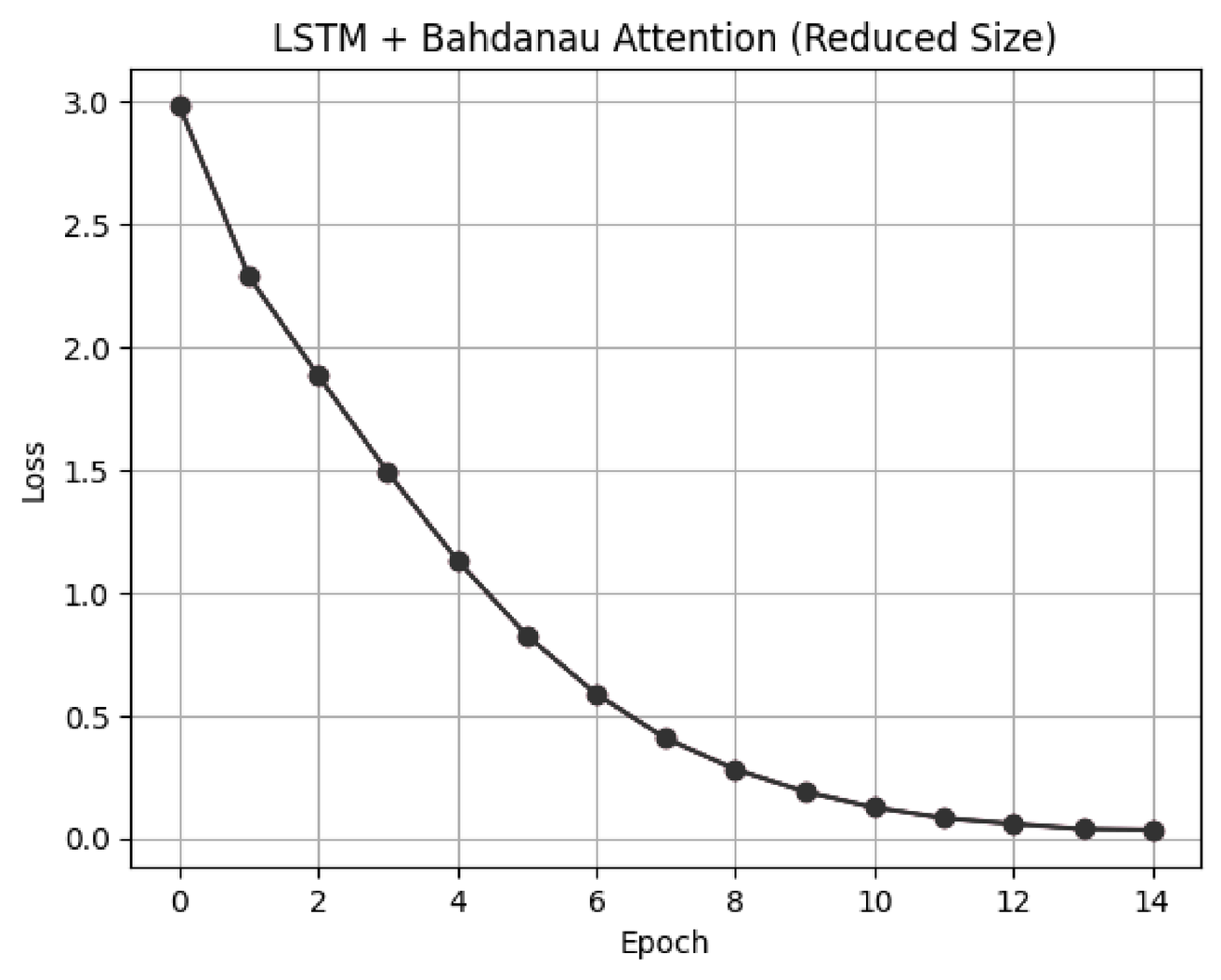
EVALUTION METRICS COMPARSIONS



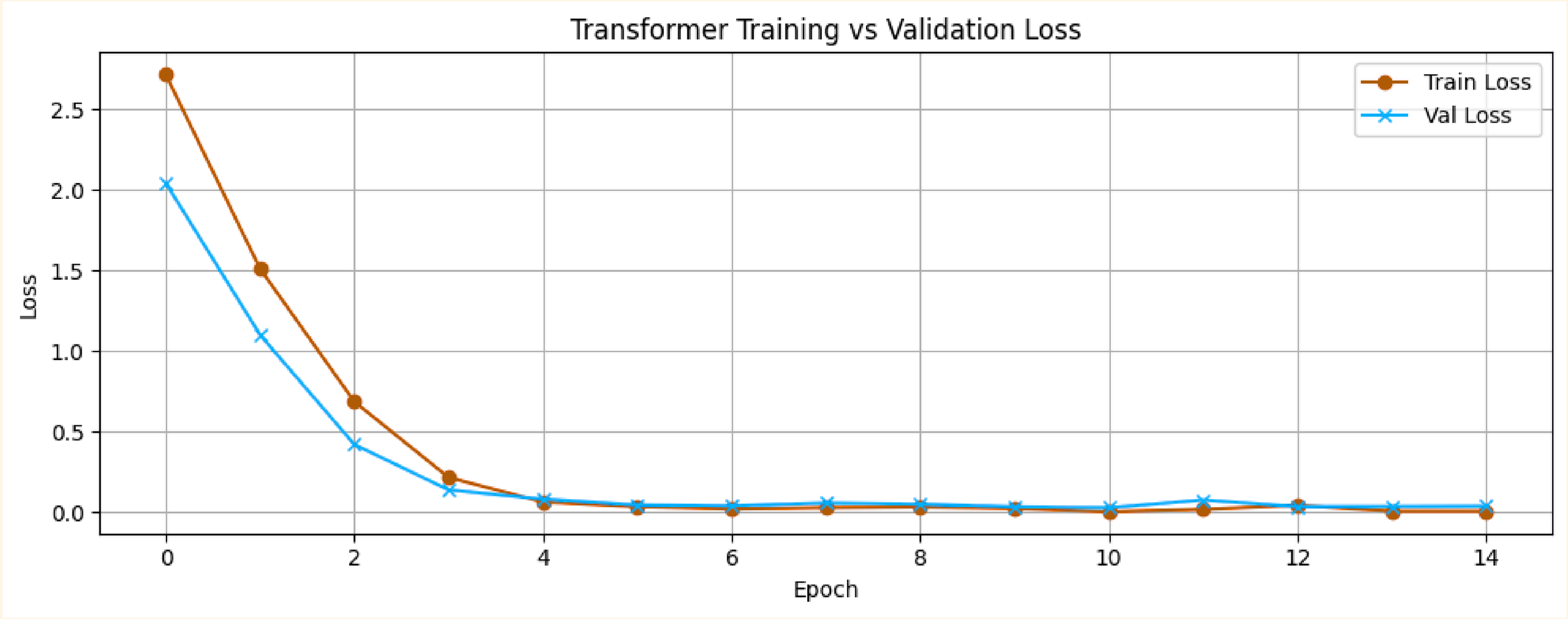
GRAPH (WITHOUT ATTENTION)



GRAPH (WITH ATTENTION)



GRAPH (SELF ATTENTION)



Conclusion

The comparative study demonstrates that the Transformer model, with its self-attention mechanism, significantly outperforms both LSTM and Bahdanau attention models in terms of grammar correction accuracy and processing efficiency. Its superior evaluation scores and faster inference time make it the most suitable choice for real-time grammar correction applications.

THANKS!