MIT Academy of Engineering
(An Autonomous Institute)

Group Members:
Mohit Muley (202201040192)
Vipul Lavhade (202201060019)
Anom Nandagawali (202201060049)

Guided by: Dr. Diptee Ghusse
Dr. Sunita Barve

Comparative Study of Encoder-Decoder Architectures for Headline Generation

Abstract

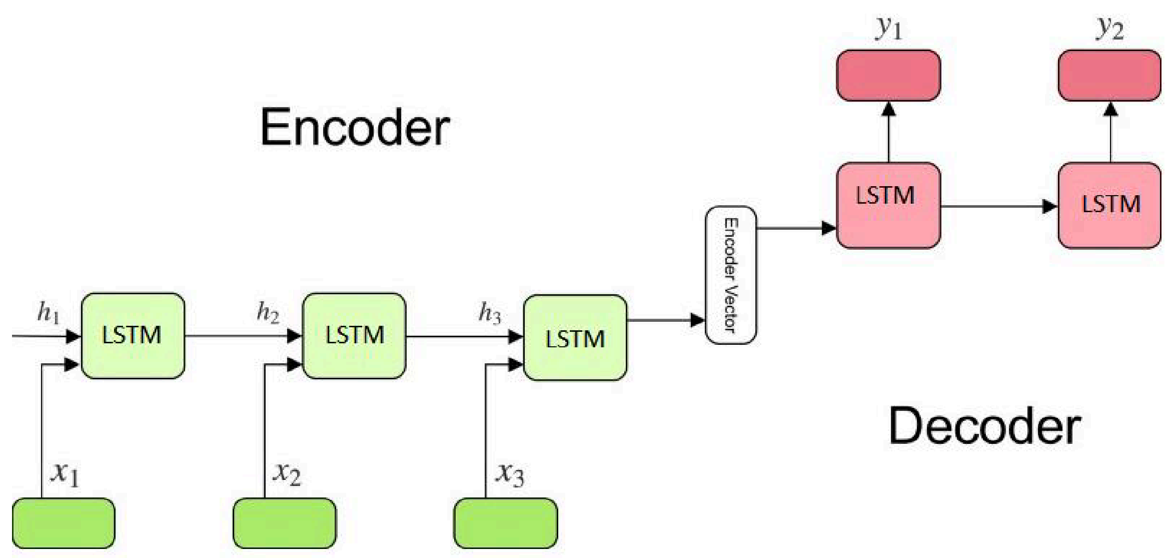
This study compares two encoder-decoder models for headline generation: LSTM/GRU without attention, and Bahdanau/Luong attention. The LSTM/GRU model serves as a baseline, while the attention models focus on key input parts, and the Transformer captures long-range dependencies through self-attention. Evaluated using BLEU, ROUGE, and METEOR metrics, the results show that attention mechanisms, especially self-attention in the Transformer, improve headline quality and efficiency.

Dataset

- We used a dataset of news articles paired with corresponding headlines.
- Preprocessing included tokenization, lowercasing, and padding to ensure consistency.
- Example:
- Article: "NASA launches new satellite to monitor climate change."
- Headline: "NASA Launches Climate Satellite"

Introduction

This project compares three encoder-decoder models for English grammar correction:
LSTM/GRU and Bahdanau Attention.
We evaluate how attention mechanisms affect correction quality and identify the most effective model.



without attention architecture

METHODOLOGY

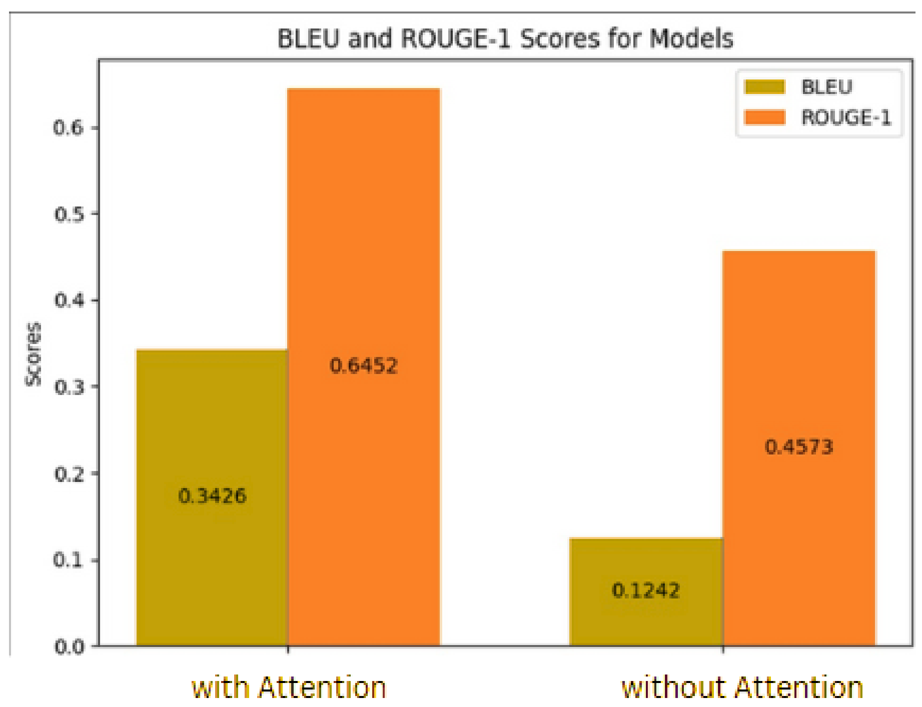
- We implemented and compared three models for headline generation:
- LSTM/GRU (No Attention): Baseline encoder-decoder model.
- Bahdanau/Luong Attention: Allows the decoder to focus on relevant input tokens during generation.
- All models were trained on article-headline pairs and evaluated using BLEU and ROUGE scores to assess summary quality.

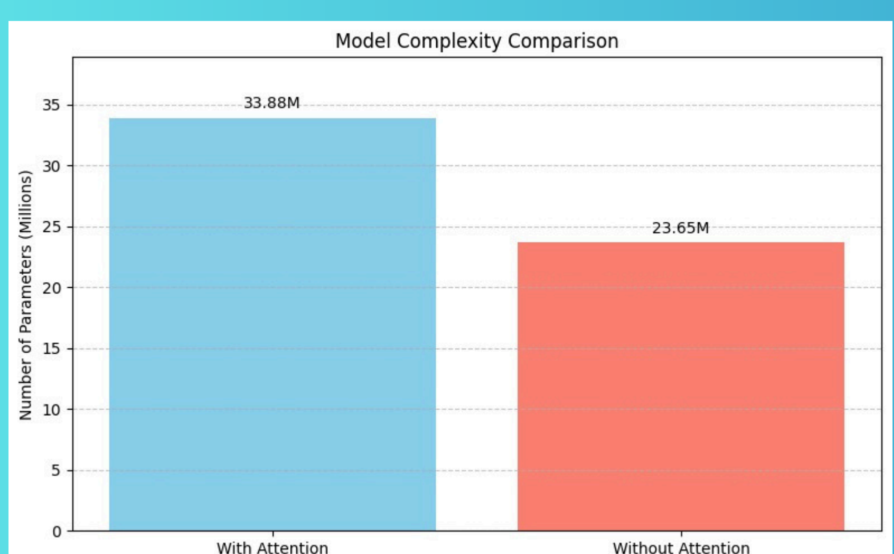
RESULTS

Model	BLEU	ROUGE-1	TRAINING TIME (SEC)	PARAM COUNT (M)
LSTM	0.124	0.4573	1196.70	23.65
ATTENTION	0.3426	0.6452	1142	33.88

Analysis

- Performance: The Attention-based model significantly outperforms the plain LSTM, achieving a BLEU score of 0.3426 and ROUGE-1 of 0.6452, compared to 0.124 and 0.4573 for LSTM.
- Efficiency: Despite having more parameters (33.88M vs. 23.65M), the Attention model trains slightly faster (1142s vs. 1196.7s).
- Insight: Incorporating attention substantially improves headline generation quality without increasing training time.





Conclusion

Attention-based models outperform those without attention in headline generation, improving both accuracy and efficiency. Future work will explore larger datasets and advanced attention architectures like Transformers for further enhancement

Research Paper Reference

1. Fact-Preserved Personalized News Headline Generation
2. Diverse Headline Generation using Self-Attention based Keyword Selection