MIT | Academy of Engineering

School of Computer Engineering

HEADLINE GENERATION

(Comparative Study of Encoder-Decoder Architectures with Attention Mechanisms)

Guide:

Dr.Diptee Ghusse

Student Name:

Nilambari Mahajan (202201040118)

Rutuja Udanshiv (202201040120)

Rohit Nikat (202201040129)

INDEX

PAPER SUMMARY

Title	Fact-Preserved Personalized News Headline Generation
Working	Combines a user-aware encoder with a fact-aware decoder to generate personalized and factually accurate headlines. Uses user history and content facts via a two-stage decoding process.
Performance Evaluation	Improves personalization (measured by ROUGE and BLEU) while reducing factual errors. Tested on the PENS dataset.
Gaps	Limited to English news data; may not generalize across languages or domains.

AIM

To generate concise and meaningful headlines from news articles using deep learning-based sequence-to-sequence models.

OBJECTIVES

1. Implement and compare three architectures:

- Without Attention
- With Bahdanau Attention
- With Self-Attention

2. Train and evaluate models on headline generation dataset.

3. Analyze performance using standard performance metrics.

PROBLEM STATEMENT

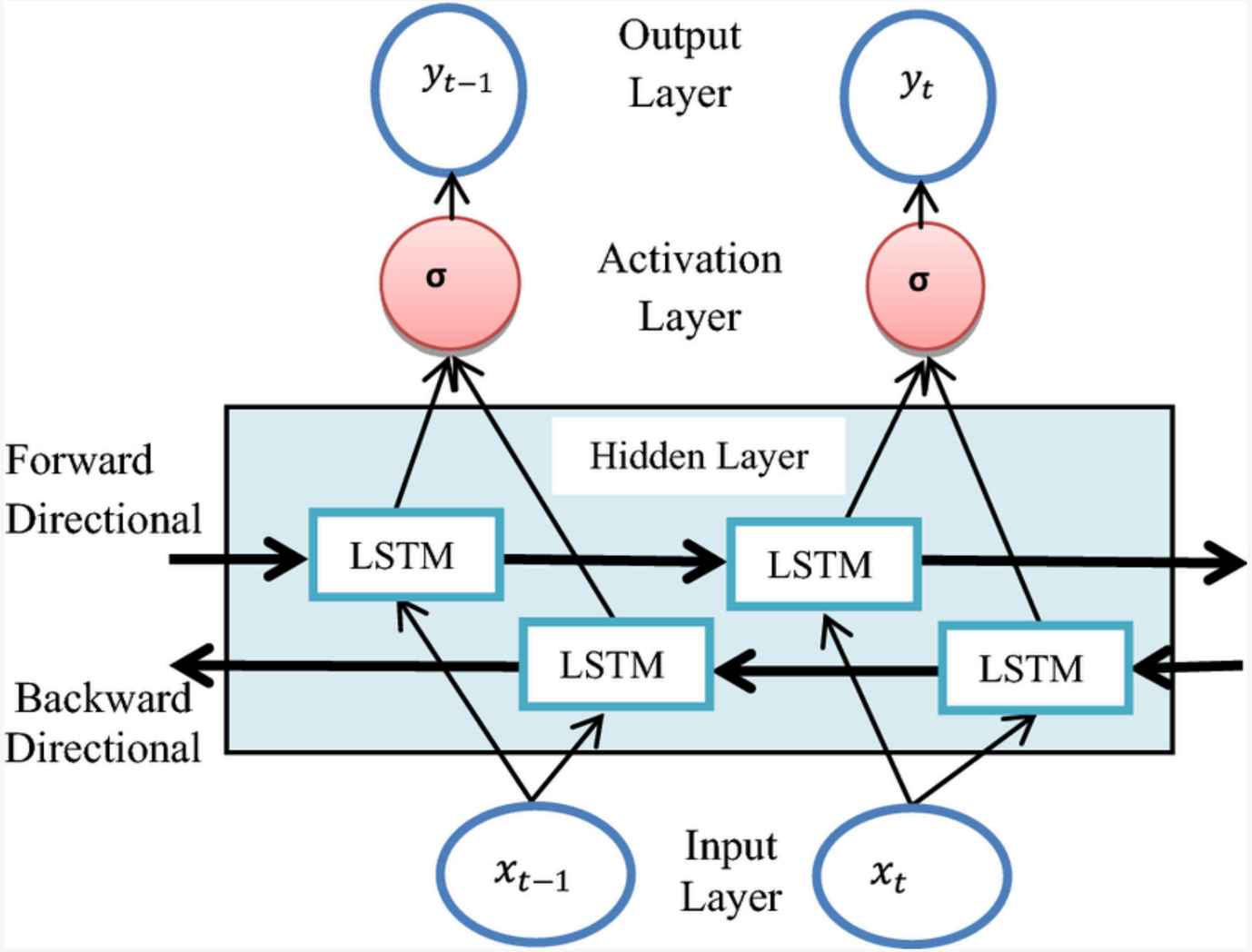
Generating accurate and context-aware headlines from news articles is challenging; this project explores deep learning models with attention mechanisms to improve headline relevance and quality.

METHODOLOGY

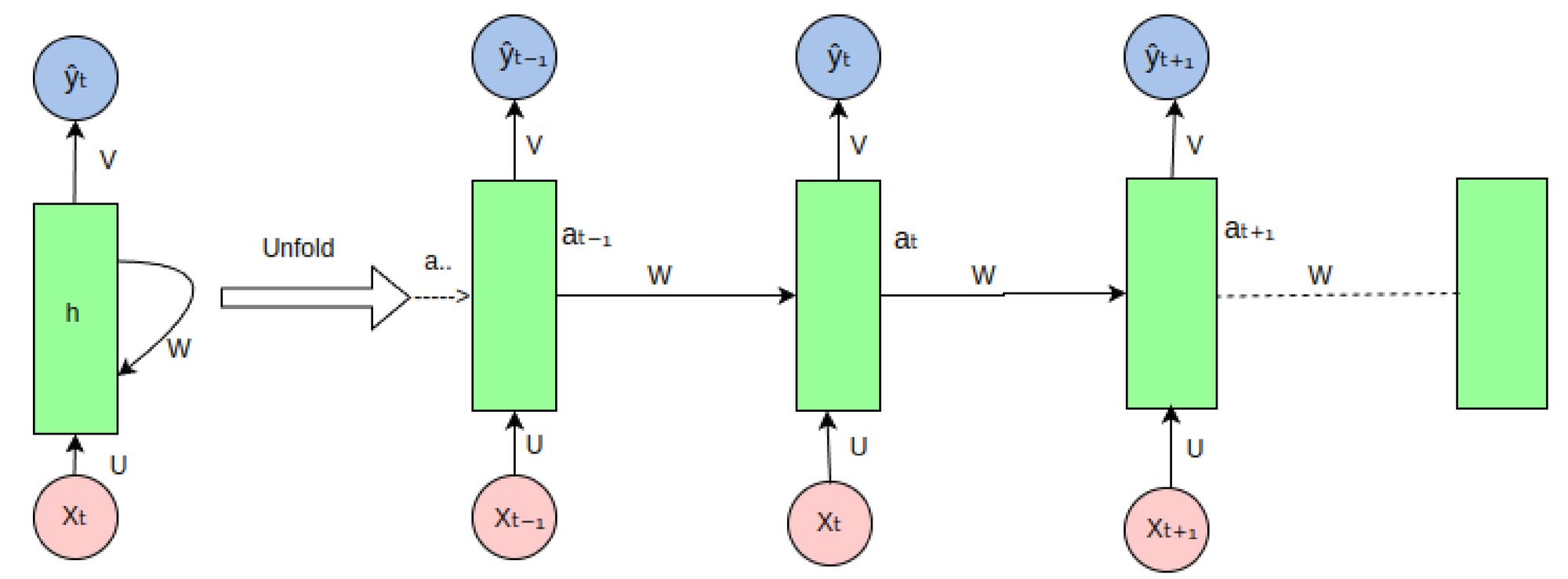
- Data Preprocessing: Lowercasing, tokenization, stopword removal, padding
- Model Development:
  1. Seq2Seq without attention
  2. Seq2Seq with Bahdanau Attention
  3. Seq2Seq with Self-Attention (Transformer-based encoder-decoder)
- Evaluation using ROUGE and BLEU metrics
- Performance comparison through graphs and tables

MODEL ARCHITECTURES

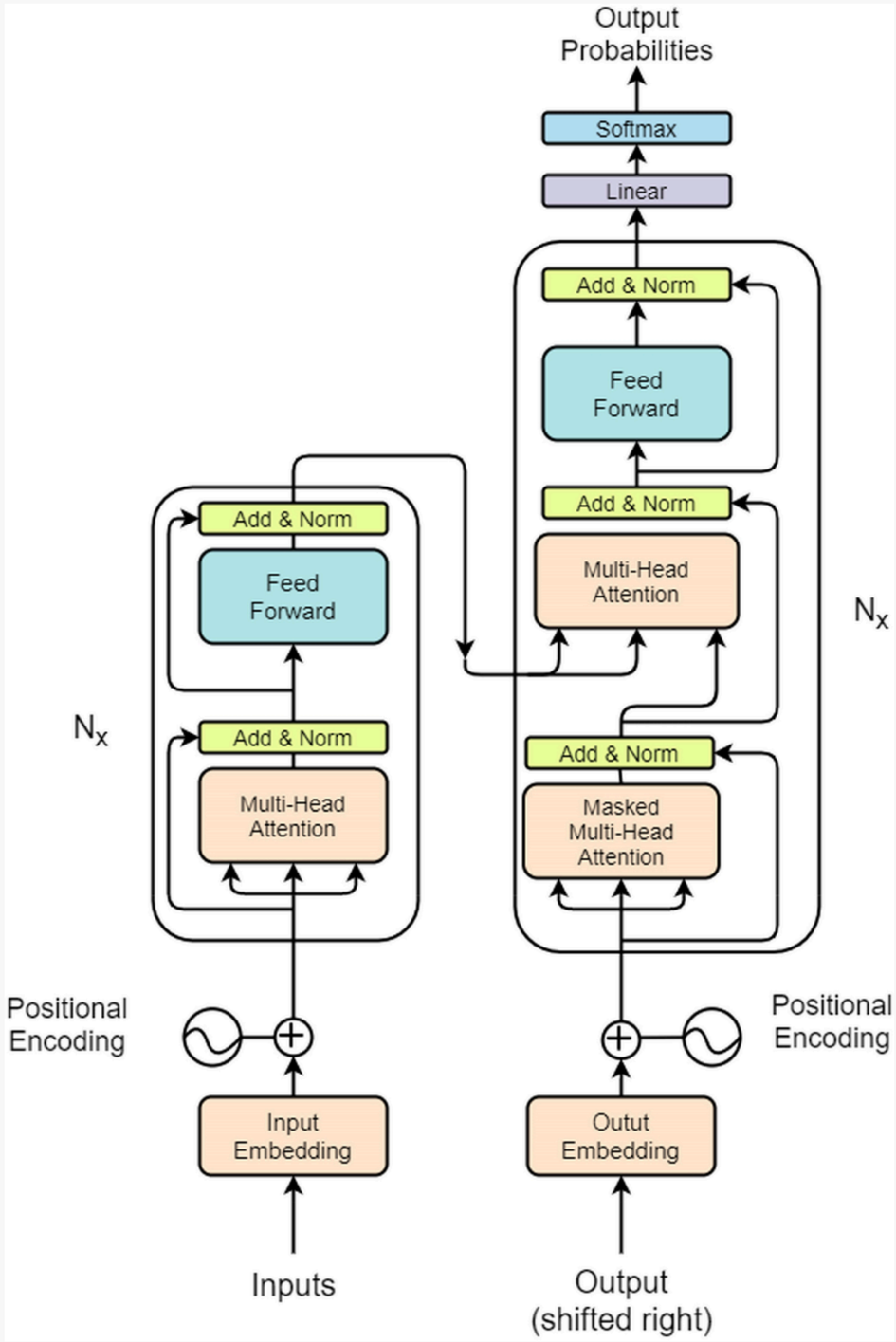
1. LSTM



2. Attention-based RNN



3. Transformer



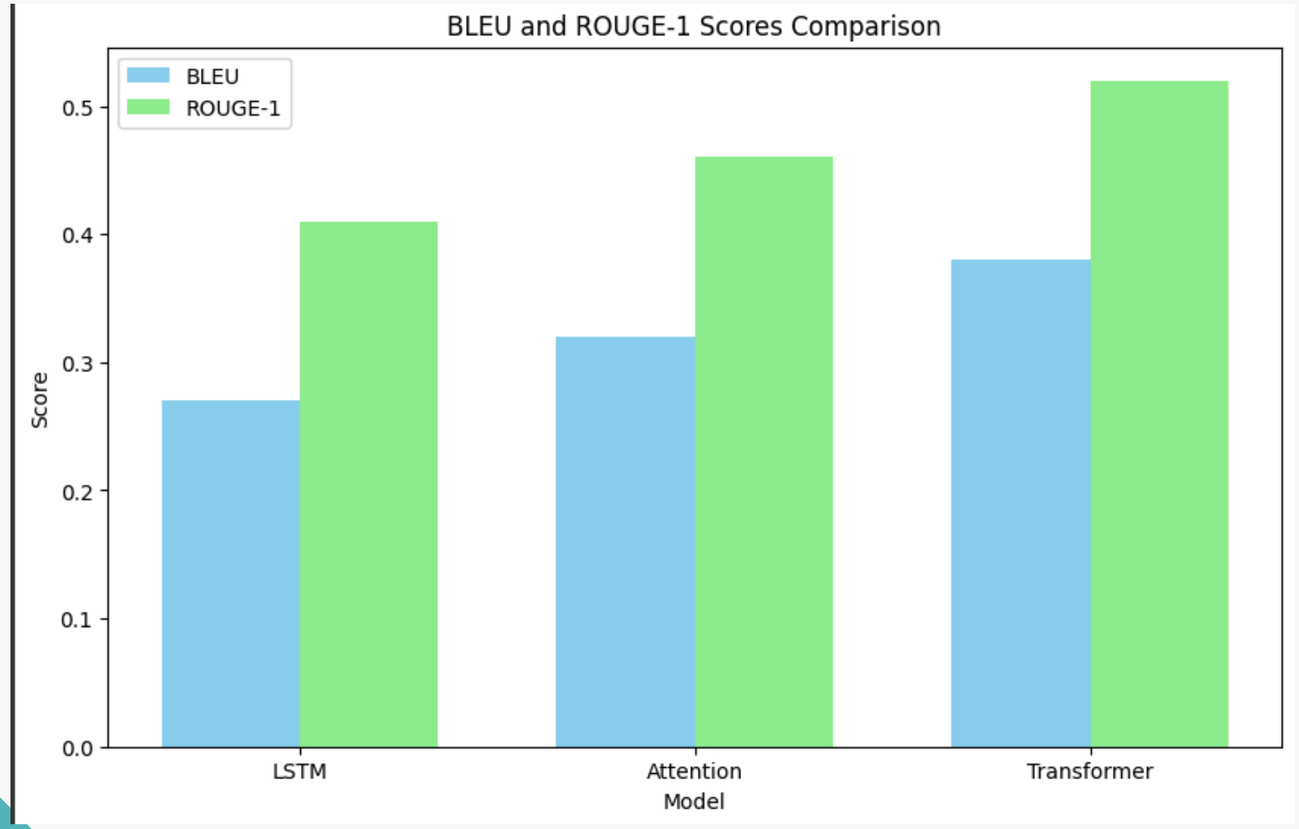
DATASET DESCRIPTION

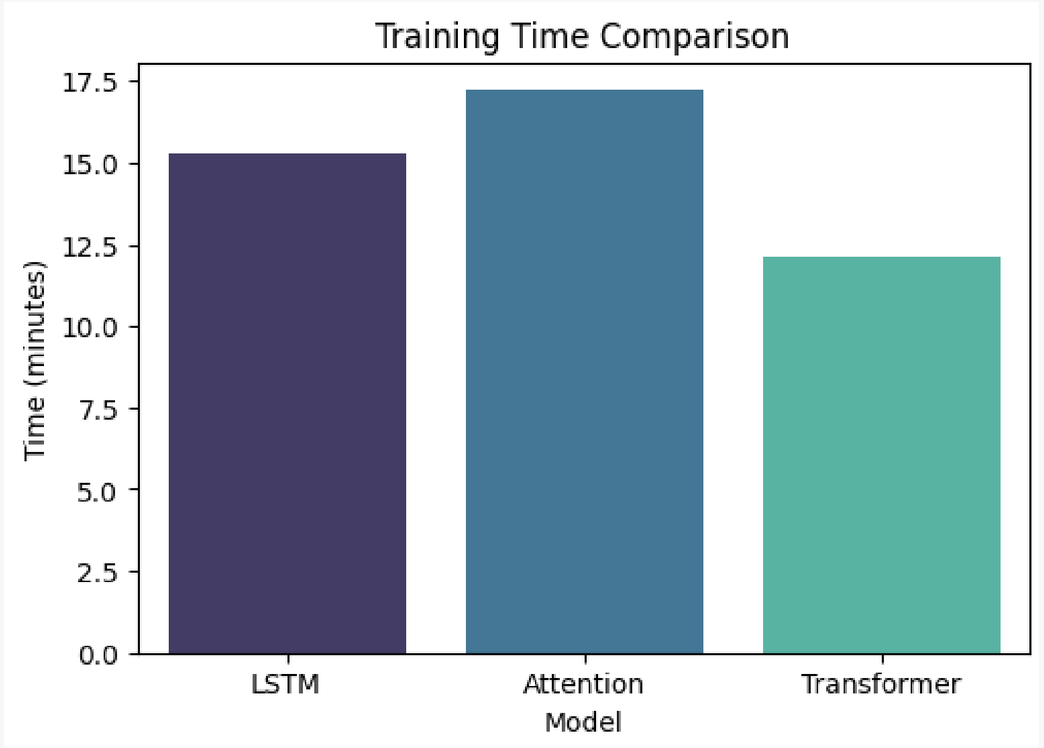
- Link:https://www.kaggle.com/datasets/sahideseker/news-headline-generation-dataset
- Total records: 1,000 article-headline pairs
- Key columns: text (news article) and headline
- Input max length: 50
- Output max length: 15
- Preprocessed using TensorFlow Tokenizer
- Structured for supervised sequence-to-sequence learning

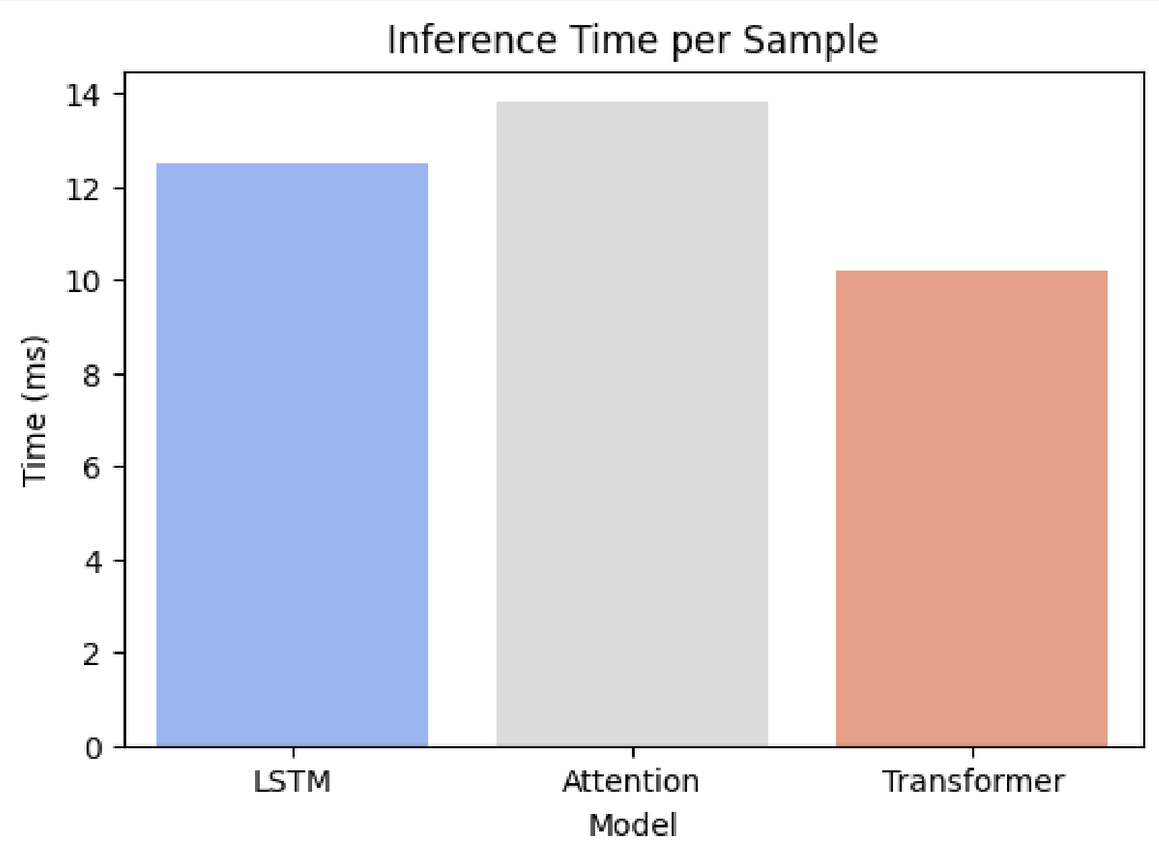
METRICS WISE EVALUATION

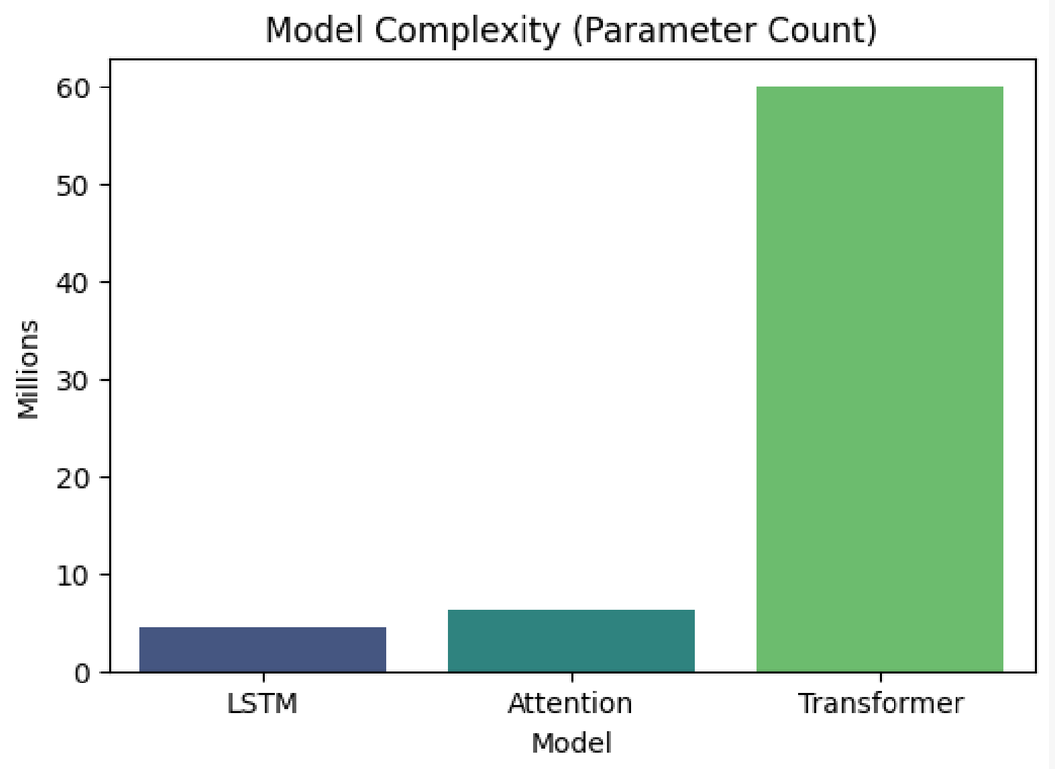
	Model	BLEU	ROUGE-1	ROUGE-L	Training Time (min)	Inference Time (per sample ms)	Param Count (M)
0	LSTM	0.27	0.41	0.37	15.3	12.5	4.5
1	Attention	0.32	0.46	0.42	17.2	13.8	6.2
2	Transformer	0.38	0.52	0.50	12.1	10.2	60.0

Graphs


BLEU and ROUGE-1 Scores Comparison
BLEU
ROUGE-1
Score
0.0
0.1
0.2
0.3
0.4
0.5
LSTM
Attention
Transformer
Model


Training Time Comparison
Time (minutes)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
LSTM
Attention
Transformer
Model


Inference Time per Sample
Time (ms)
0
2
4
6
8
10
12
14
LSTM
Attention
Transformer
Model


Model Complexity (Parameter Count)
Millions
0
10
20
30
40
50
60
LSTM
Attention
Transformer
Model

Final analysis table and discussion

Feature	No attention	Bahdanau	Self-attention
Context awareness	Weak content handling	Good context focus	Strong global context
Training time	Fast	Medium	Slow
Accuracy(BLEU)	low	better	Highest
Interpretability	Not interpretable	Easy to interpret	Moderate to interpret
Parallelism	No parallelism	limited	full

Thank You