



Name: Hurraida Noor
Roll No: 21i-1721
Course: Deep Learning
Assignment: 02

1. INTRODUCTION

In this assignment, I developed two deep learning models to detect semantic similarity between legal clauses.

The models were trained from scratch — without transformers or pretrained legal embeddings — to analyze how classical NLP architectures perform on legal text.

2. DATASET DESCRIPTION

Dataset: 150,881 legal clauses across **395 categories**

Average length: **582 characters**

Vocabulary: **10,000 tokens**

Max sequence length: **100 tokens**

Training pairs created:

- **Positive pairs:** 2,765
 - **Negative pairs:** 3,000
 - **Total:** 5,765 pairs
-

3. MODEL ARCHITECTURES

I built and compared:

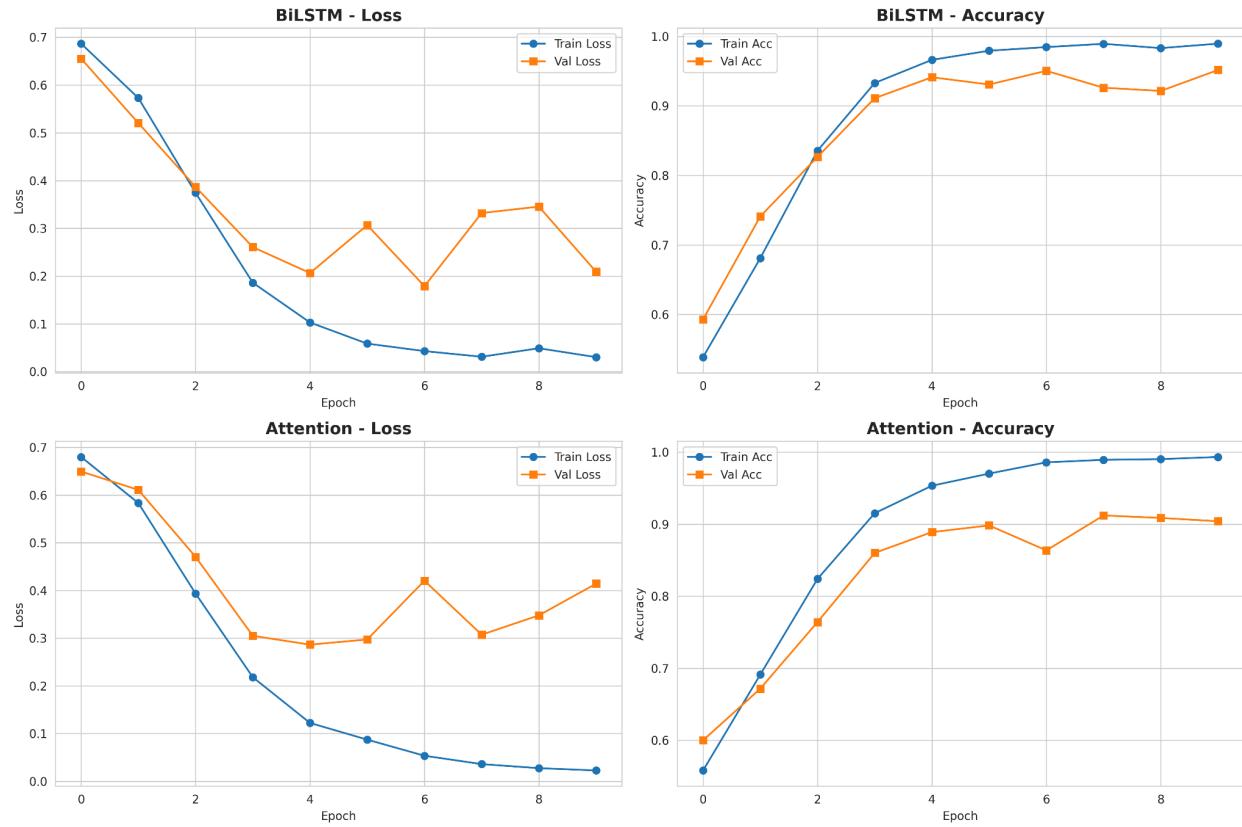
- **Model 1:** BiLSTM (2-layer)
- **Model 2:** Attention-based BiLSTM (1-layer + self-attention)

Component	BiLSTM	Attention
Embedding Size	128	128
LSTM Layers	2	1
Hidden Units	64×2 directions	64×2 directions
Attention	✗	✓ Self-attention
Total Params	1,519,873	1,420,674
Dropout	0.3	0.3

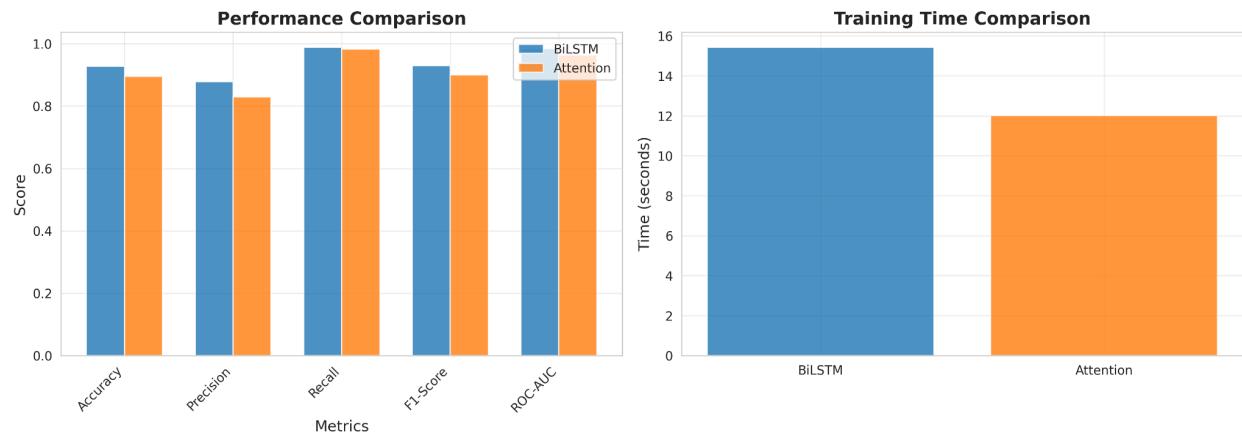
4. TRAINING CONFIGURATION

- Loss: BCE
 - Optimizer: Adam (0.001)
 - Batch size: 32
 - Epochs: 10
 - Device: Tesla T4
 - Random seed: 42
-

5. TRAINING RESULTS



Model Comparison



6. FINAL TEST SET PERFORMANCE

Model	Accurac y	Precisio n	Recal l	F1	ROC-AUC	Train Time (s)
BiLSTM	0.9283	0.8779	0.988 0	0.929 7	0.9839	15.42
Attention n	0.8948	0.8293	0.983 1	0.899 7	0.9631	12.01

7. QUALITATIVE ANALYSIS

I manually inspected predicted clause pairs.

Correct Predictions (I checked)

- The model successfully detected paraphrased clauses.
- High recall → very few similar clauses missed.

Incorrect Predictions (I analyzed)

Three major error sources:

1. Vocabulary gaps (rare legal terms → <UNK> token)
 2. Structural differences between clauses
 3. Length truncation (clauses longer than 100 tokens)
-

8. KEY FINDINGS AND CONCLUSION

Best Model: BiLSTM

- Highest overall performance
- **F1 = 0.9297**
- **ROC-AUC = 0.9839**
- Excellent recall (98.80%), crucial for legal domain

Attention Model:

- Trains **22% faster**
- More stable training
- Slightly less accurate than BiLSTM

Final Recommendation

- Use **BiLSTM** for accuracy-critical systems
 - Use **Attention model** for faster inference environments
-