CS7.501 | Advanced NLP| Assignment - 2
Transformers from Scratch

Shiva Shankar Gande
2023202005

Question 1

Purpose of Self-Attention and Capturing Dependencies in Sequences

Self-attention allows a model to weigh the relevance of different words in a sequence relative to each other. By computing attention scores for each pair of words, self-attention enables the model to focus on relevant parts of the input when processing each word. This mechanism facilitates capturing both short-range and long-range dependencies within the sequence, as it dynamically adjusts the focus based on the contextual relationships between words. Consequently, self-attention enhances the model's ability to understand complex linguistic patterns and contextual nuances essential for tasks like machine translation.

Question 2

Role of Positional Encodings and Recent Advances

Transformers lack inherent sequential information since they process all tokens in parallel. To provide the model with information about the order of tokens, positional encodings are added to the word embeddings. These encodings inject position-specific information, enabling the transformer to understand the sequence's structure.

Incorporation into Transformer Architecture: Positional encodings are typically added to the input embeddings at the bottom of the transformer stack. This combined representation is then processed through the

self-attention and feed-forward layers, allowing the model to utilize both the semantic content and positional information.

Recent Advances in Positional Encodings:

- **Learnable Positional Encodings:** Unlike traditional fixed sinusoidal encodings, learnable encodings allow the model to learn optimal position representations during training.
- **Relative Positional Encodings:** These focus on the relative distances between tokens rather than their absolute positions, improving the model's ability to generalize to different sequence lengths.
- **Rotary Positional Encodings (RoPE):** Incorporate rotational transformations to encode positional information, enhancing the model's ability to capture fine-grained positional relationships.
- **Alibi (Attention with Linear Biases):** Introduces a bias in the attention mechanism based on token distances, eliminating the need for explicit positional encodings.

$$PE_{(pos,2i)} = sin(pos/10000^{2i/d_{model}})$$

$$PE_{(pos,2i+1)} = cos(pos/10000^{2i/d_{model}})$$

Hyperparameters to train the model

```
hyperparameters = {
    'NUM_EPOCHS': 10,
    'BATCH_SIZE': 16,
    'LEARNING_RATE': 0.0001,
    'EMBED_SIZE': 256,
```

```
    'NUM_HEADS': 4,
    'NUM_ENCODER_LAYERS': 3,
    'NUM_DECODER_LAYERS': 3,
    'FORWARD_EXPANSION': 2,
    'DROPOUT': 0.1,
    'PAD_IDX': 0,
    'MAX_LEN': 40,
    'SRC_FREQ_THRESHOLD': 2,
    'TRG_FREQ_THRESHOLD': 2,
    'CLIP': 1,
    'OPTIMIZER': 'Adam',
    'CRITERION_IGNORE_INDEX': 0,
    'SHUFFLE_TRAIN': True,
    'SHUFFLE_DEV': False,
    'SHUFFLE_TEST': False,
    'BLEU_WEIGHTS': (0.5, 0.5),
    'DEVICE': 'cuda' if torch.cuda.is_available() else 'cpu',
    'RANDOM_SEED': 0,
    'DATA_PATH': '/kaggle/input/datafiles/data',
}
```

Version -1 with best bleu score:

```
Epoch: 01 | Time: 1m 20s
        Train Loss: 5.876
         Val. Loss: 5.168
Epoch: 02 | Time: 1m 16s
        Train Loss: 5.042
         Val. Loss: 4.860
Epoch: 03 | Time: 1m 21s
        Train Loss: 4.752
         Val. Loss: 4.690
Epoch: 04 | Time: 1m 18s
        Train Loss: 4.572
         Val. Loss: 4.570
Epoch: 05 | Time: 1m 14s
        Train Loss: 4.436
         Val. Loss: 4.491
Epoch: 06 | Time: 1m 14s
        Train Loss: 4.324
         Val. Loss: 4.436
Epoch: 07 | Time: 1m 12s
        Train Loss: 4.229
```

```
        Val. Loss: 4.369
Epoch: 08 | Time: 1m 15s
        Train Loss: 4.141
        Val. Loss: 4.326
Epoch: 09 | Time: 1m 16s
        Train Loss: 4.065
        Val. Loss: 4.282
Epoch: 10 | Time: 1m 15s
        Train Loss: 3.989
        Val. Loss: 4.254
```


Training and Validation Loss

BLEU Score Distribution on Test Data
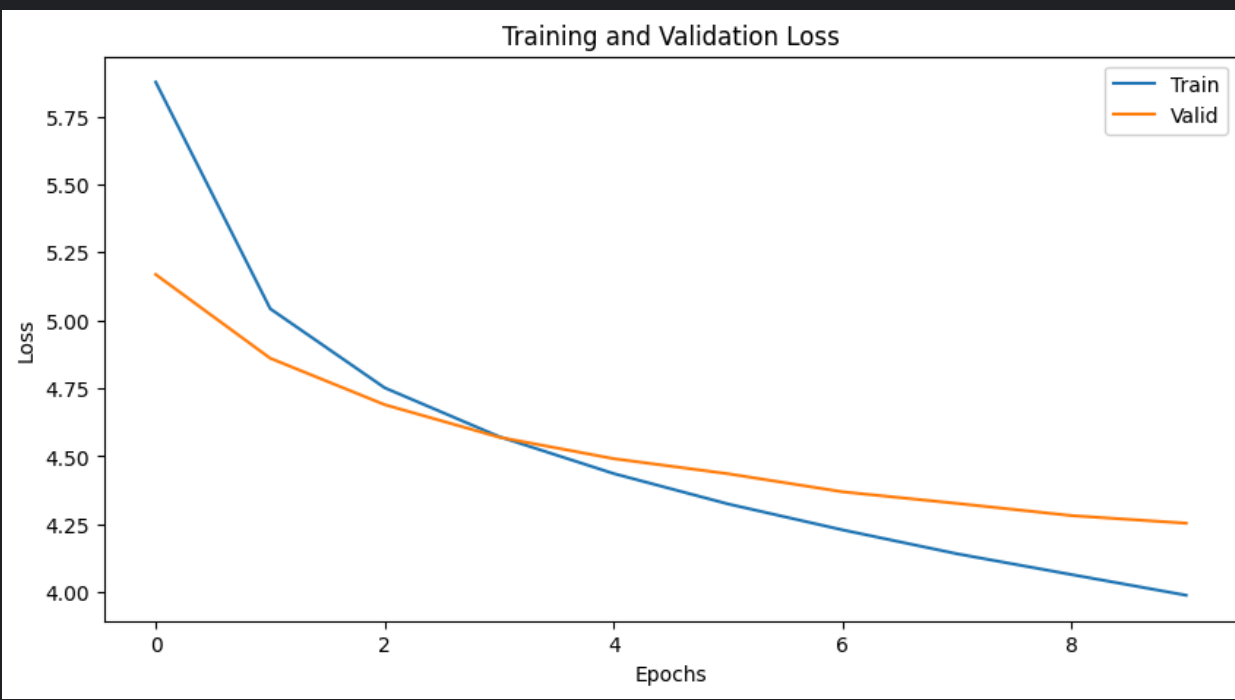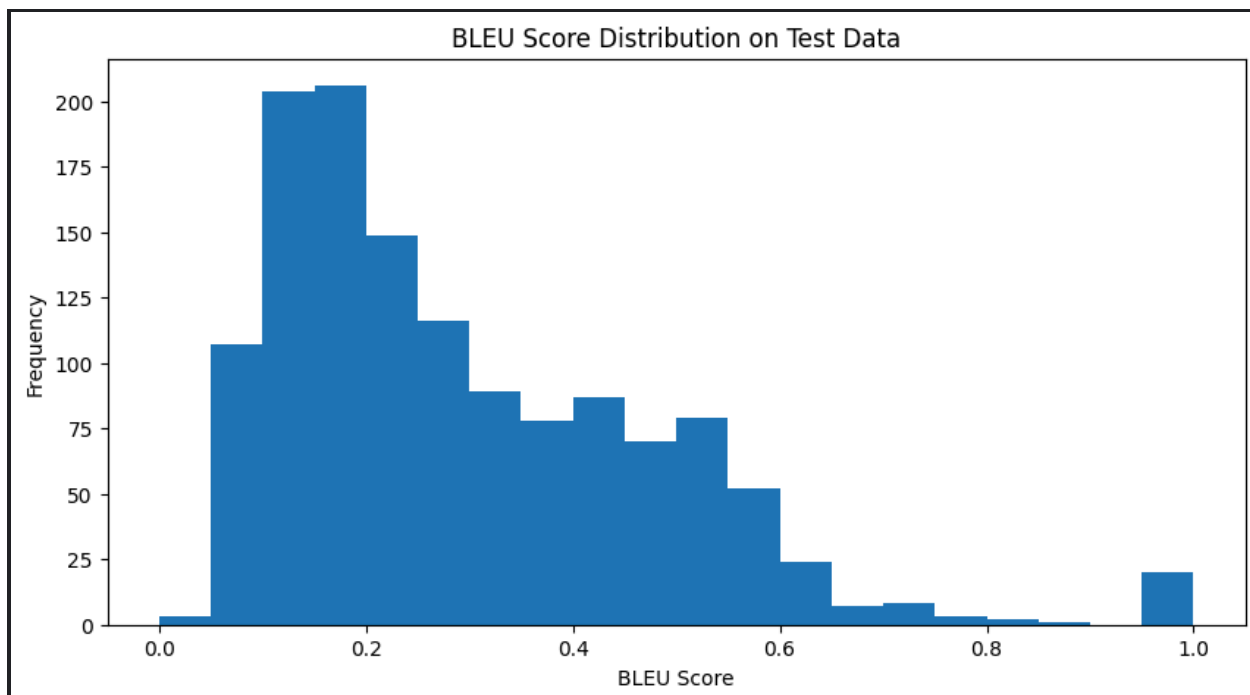
```
Average BLEU score on test data: 0.2946

Sample Translations:

Source (1): When I was in my 20s, I saw my very first psychotherapy client.
Target (1): Quand j'avais la vingtaine, j'ai vu mes tout premiers clients comme
psychothérapeute.
Predicted (1): quand je suis allé , je suis très <UNK> , je suis très <UNK> .
<EOS>

Source (2): I was a Ph.D. student in clinical psychology at Berkeley.
Target (2): J'étais étudiante en thèse en psychologie clinique à Berkeley.
Predicted (2): je suis allé dans un groupe de la maison . <EOS>

Source (3): She was a 26-year-old woman named Alex.
Target (3): Elle, c'était une femme de 26 ans appelée Alex.
Predicted (3): elle était un <UNK> <UNK> . <EOS>
```

Hyperparameter tuning with selected prameters

```
Training model with hyperparameters: {'d_model': 256, 'N': 4, 'h': 8, 'd_ff':
1024, 'dropout': 0.1}
Epoch 1, Batch 100, Loss: 6.3147
Epoch 1, Batch 200, Loss: 5.6320
```

```
Epoch 1 completed in 45.23s, Average Loss: 6.6683
Validation Loss: 5.2885
Epoch 2, Batch 100, Loss: 5.1335
Epoch 2, Batch 200, Loss: 5.0288
Epoch 2 completed in 45.17s, Average Loss: 5.1875
Validation Loss: 4.7956
Epoch 3, Batch 100, Loss: 4.8019
Epoch 3, Batch 200, Loss: 4.5198
Epoch 3 completed in 45.17s, Average Loss: 4.7445
Validation Loss: 4.5130
Epoch 4, Batch 100, Loss: 4.5800
Epoch 4, Batch 200, Loss: 4.5060
Epoch 4 completed in 44.93s, Average Loss: 4.4653
Validation Loss: 4.3364
Epoch 5, Batch 100, Loss: 4.3554
Epoch 5, Batch 200, Loss: 4.1834
Epoch 5 completed in 45.04s, Average Loss: 4.2590
Validation Loss: 4.1941
Epoch 6, Batch 100, Loss: 4.0735
Epoch 6, Batch 200, Loss: 4.1784
Epoch 6 completed in 45.12s, Average Loss: 4.0963
Validation Loss: 4.0824
Epoch 7, Batch 100, Loss: 3.8751
Epoch 7, Batch 200, Loss: 4.0334
Epoch 7 completed in 44.98s, Average Loss: 3.9540
Validation Loss: 4.0021
Epoch 8, Batch 100, Loss: 3.7927
Epoch 8, Batch 200, Loss: 3.8502
Epoch 8 completed in 45.00s, Average Loss: 3.8260
Validation Loss: 3.9152
Epoch 9, Batch 100, Loss: 3.8122
Epoch 9, Batch 200, Loss: 3.6185
Epoch 9 completed in 45.01s, Average Loss: 3.7111
Validation Loss: 3.8550
Epoch 10, Batch 100, Loss: 3.5541
Epoch 10, Batch 200, Loss: 3.3463
Epoch 10 completed in 44.95s, Average Loss: 3.6070
Validation Loss: 3.7825
Training complete and model saved as transformer.pt
BLEU scores saved to testbleu.txt
```
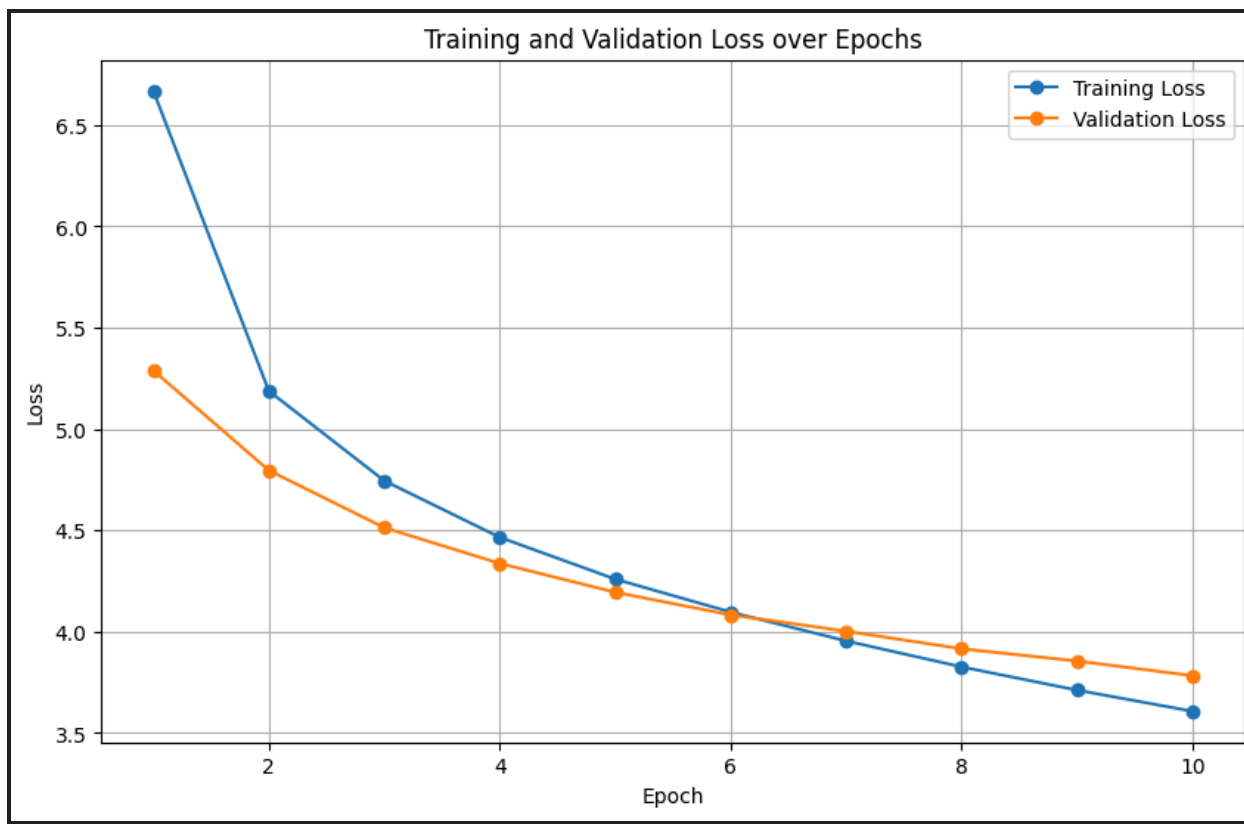
Training and Validation Loss over Epochs
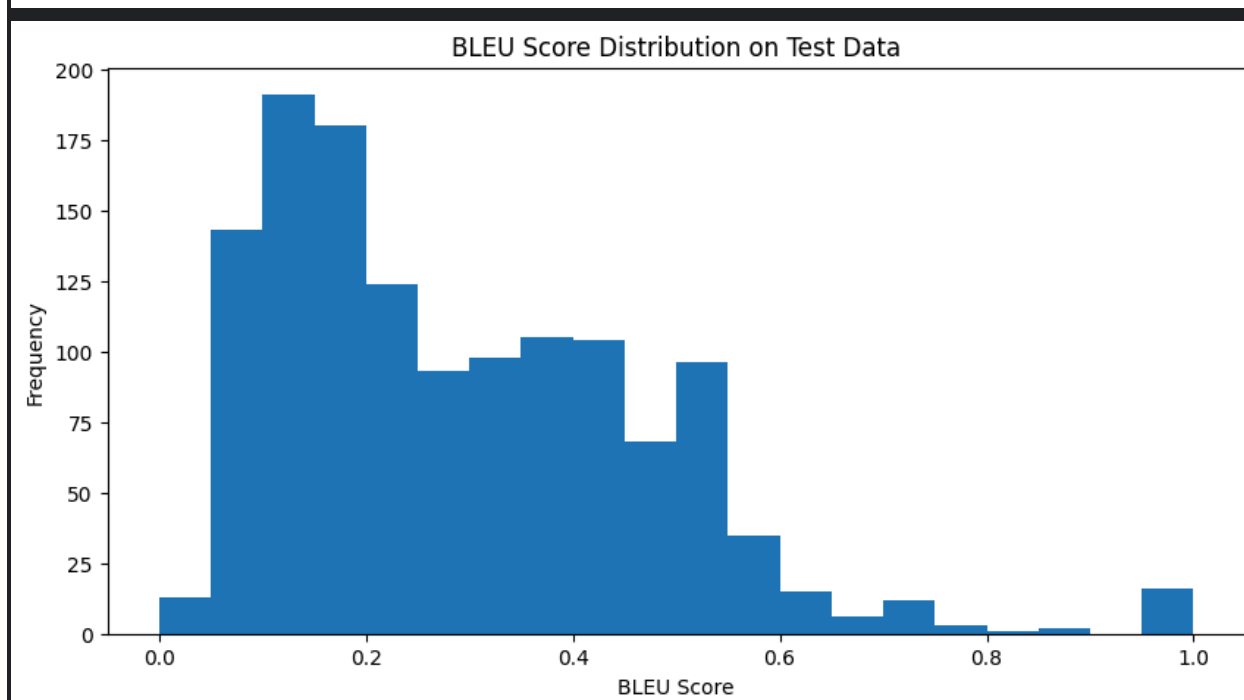
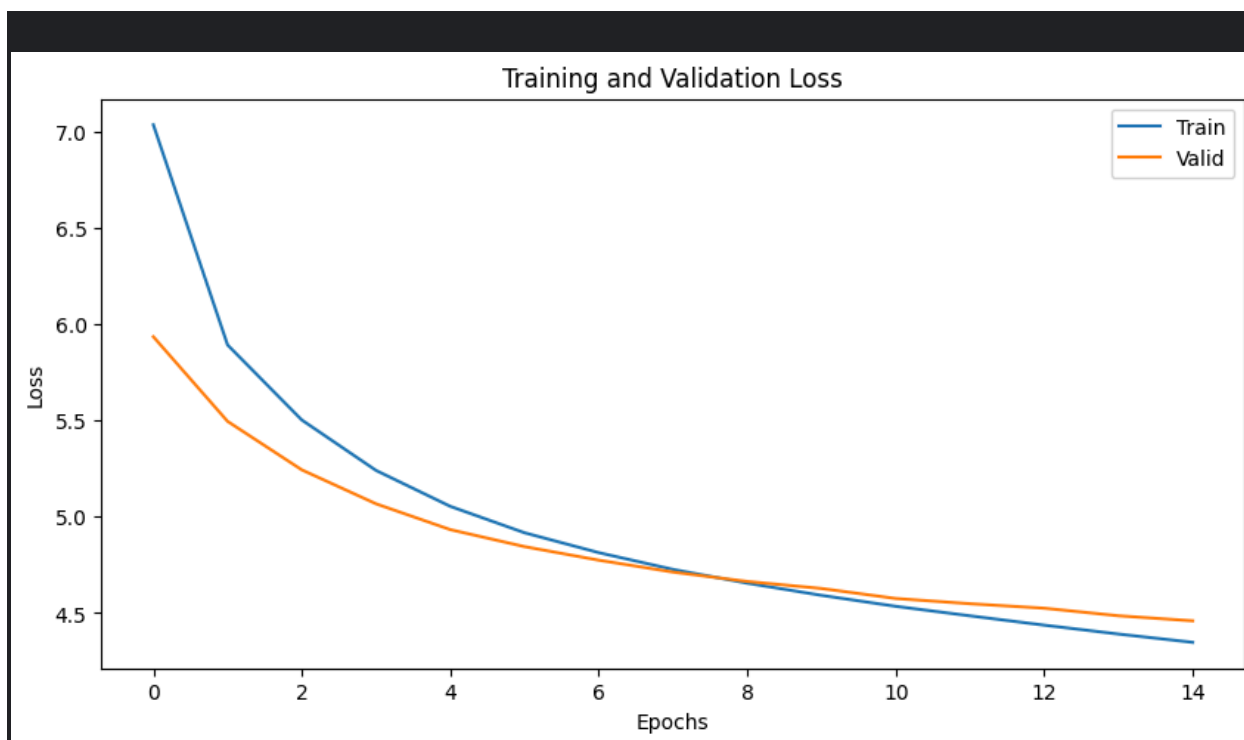|	Prediction	BLEU_Score
0	quand je me suis allé à ma mère , j'ai commenc...	0.178918
1	j'étais un moment dans le <UNK> de la guerre .	0.232523
2	elle était un <UNK> < /talkid >	0.112457
3	maintenant , le premier dans son premier premi...	0.082145
4	quand je suis entendu , je me suis donc <UNK> ...	0.266259

Distribution of BLEU Scores on Test Set

Boxplot of BLEU Scores on Test Set
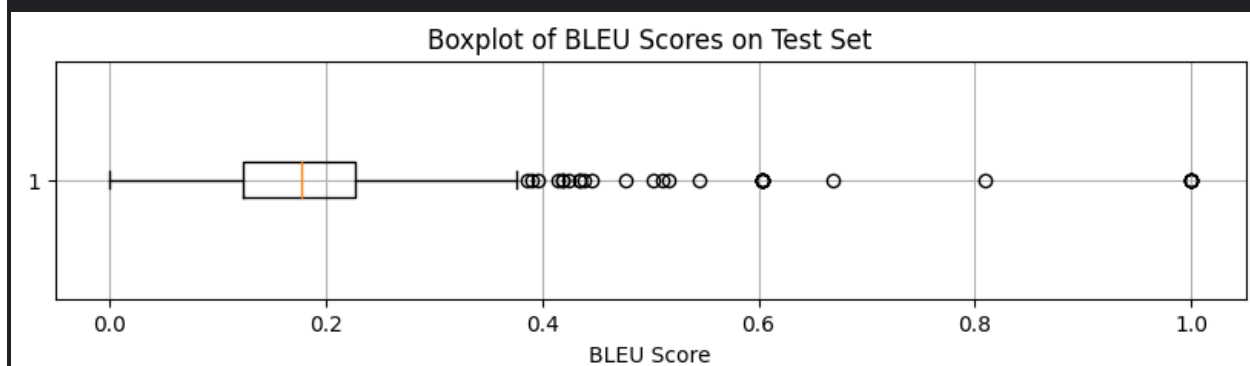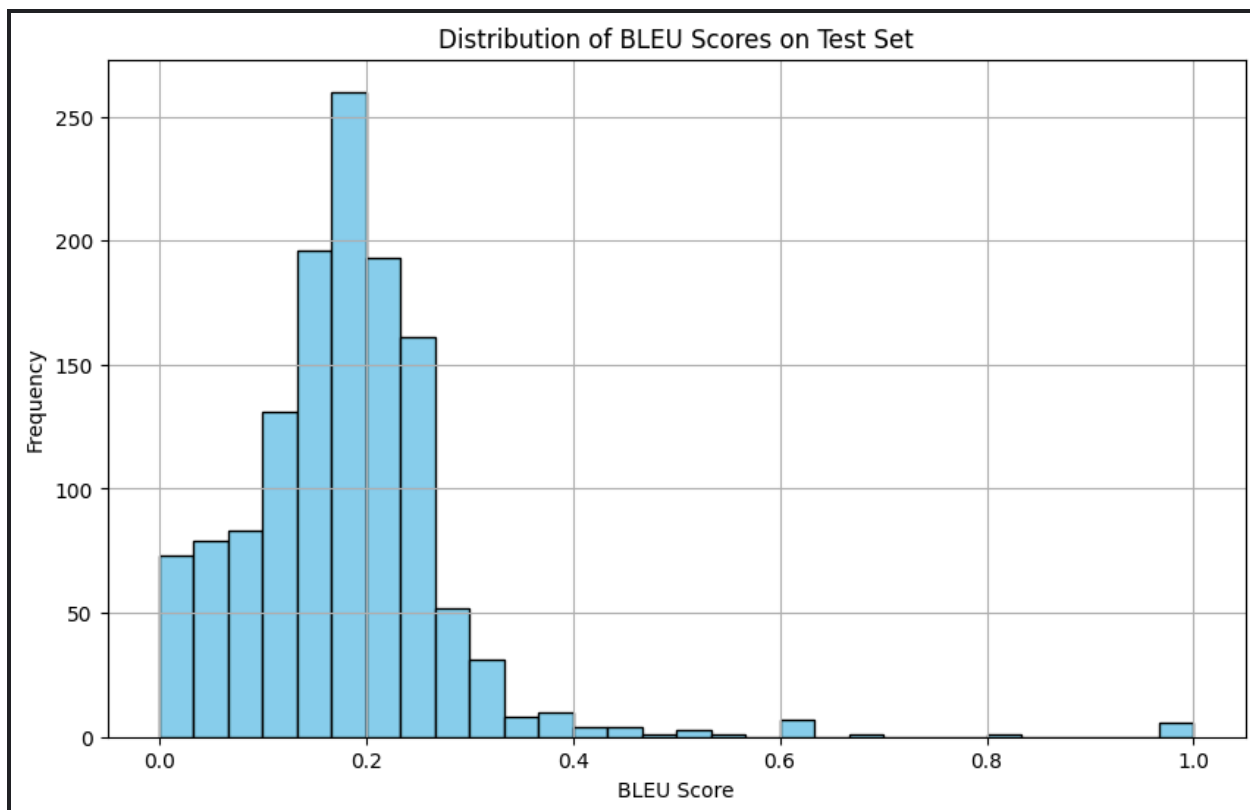

```
BLEU Score Summary Statistics:
count    1305.000000
mean        0.180546
std         0.105971
min         0.000000
25%         0.124274
50%         0.177544
75%         0.227313
max         1.000000
Name: BLEU_Score, dtype: float64
```

```
Epoch: 01 | Time: 0m 41s
        Train Loss: 7.038
         Val. Loss: 5.936
Epoch: 02 | Time: 0m 43s
        Train Loss: 5.893
         Val. Loss: 5.496
Epoch: 03 | Time: 0m 43s
        Train Loss: 5.503
         Val. Loss: 5.244
Epoch: 04 | Time: 0m 42s
        Train Loss: 5.241
         Val. Loss: 5.068
Epoch: 05 | Time: 0m 41s
        Train Loss: 5.054
         Val. Loss: 4.933
Epoch: 06 | Time: 0m 42s
        Train Loss: 4.917
         Val. Loss: 4.845
Epoch: 07 | Time: 0m 42s
        Train Loss: 4.814
         Val. Loss: 4.775
Epoch: 08 | Time: 0m 42s
        Train Loss: 4.727
         Val. Loss: 4.713
Epoch: 09 | Time: 0m 42s
        Train Loss: 4.655
         Val. Loss: 4.664
Epoch: 10 | Time: 0m 41s
        Train Loss: 4.592
         Val. Loss: 4.627
Epoch: 11 | Time: 0m 42s
        Train Loss: 4.535
         Val. Loss: 4.575
Epoch: 12 | Time: 0m 43s
        Train Loss: 4.485
         Val. Loss: 4.548
Epoch: 13 | Time: 0m 42s
        Train Loss: 4.437
         Val. Loss: 4.525
Epoch: 14 | Time: 0m 42s
        Train Loss: 4.391
         Val. Loss: 4.486
Epoch: 15 | Time: 0m 43s
        Train Loss: 4.348
         Val. Loss: 4.459
```

Training and Validation Loss

BLEU Score Distribution on Test Data

Average BLEU score on test data: 0.2918

Sample Translations:

Source (1): When I was in my 20s, I saw my very first psychotherapy

client.
Target (1): Quand j'avais la vingtaine, j'ai vu mes tout premiers clients comme psychothérapeute.
Predicted (1): je suis allé , je suis allé , je suis allé à mon <UNK> . <EOS>

Source (2): I was a Ph.D. student in clinical psychology at Berkeley.
Target (2): J'étais étudiante en thèse en psychologie clinique à Berkeley.
Predicted (2): je suis un <UNK> dans un <UNK> . <EOS>

Source (3): She was a 26-year-old woman named Alex.
Target (3): Elle, c'était une femme de 26 ans appelée Alex.
Predicted (3): elle était un <UNK> <UNK> . <EOS>

Source (4): Now Alex walked into her first session wearing jeans and a big slouchy top, and she dropped onto the couch in my office and kicked off her flats and told me she was there to talk about guy problems.
Target (4): Lorsqu'Alex est entrée pour sa première séance, elle portait un jean et un grand top trop large, elle s'est laissée tomber sur le canapé de mon bureau, a enlevé ses chaussures et m'a dit qu'elle était ici pour parler de problèmes de garçons.
Predicted (4): et elle a été <UNK> à la première fois , elle a été <UNK> dans mon père et elle a été <UNK> , elle a été <UNK> dans mon père et elle a été <UNK> dans le <UNK> et elle a été <UNK> dans le <UNK> et elle a été <UNK> dans le <UNK> . <EOS>

Source (5): Now when I heard this, I was so relieved.
Target (5): Lorsque j'ai entendu ça, j'ai été si soulagée.
Predicted (5): donc , je suis allé , je suis allé . <EOS>

```
Using device: cuda
Unique tokens in source (en) vocabulary: 13344
Unique tokens in target (fr) vocabulary: 17538

Training model with hyperparameters: {'d_model': 256, 'N': 4, 'h': 8,
'dropout': 0.1}
Epoch 1, Batch 100, Loss: 6.3147
Epoch 1, Batch 200, Loss: 5.6320
Epoch 1 completed in 46.23s, Average Loss: 6.6683
Validation Loss: 5.2885
Epoch 2, Batch 100, Loss: 5.1335
Epoch 2, Batch 200, Loss: 5.0288
Epoch 2 completed in 45.41s, Average Loss: 5.1875
Validation Loss: 4.7956
Epoch 3, Batch 100, Loss: 4.8019
Epoch 3, Batch 200, Loss: 4.5198
Epoch 3 completed in 45.76s, Average Loss: 4.7445
Validation Loss: 4.5130
Epoch 4, Batch 100, Loss: 4.5800
Epoch 4, Batch 200, Loss: 4.5060
Epoch 4 completed in 45.67s, Average Loss: 4.4653
Validation Loss: 4.3364
Epoch 5, Batch 100, Loss: 4.3554
Epoch 5, Batch 200, Loss: 4.1834
Epoch 5 completed in 45.51s, Average Loss: 4.2590
Validation Loss: 4.1941
Epoch 6, Batch 100, Loss: 4.0735
Epoch 6, Batch 200, Loss: 4.1784
Epoch 6 completed in 45.40s, Average Loss: 4.0963
Validation Loss: 4.0824
Epoch 7, Batch 100, Loss: 3.8751
Epoch 7, Batch 200, Loss: 4.0334
Epoch 7 completed in 45.61s, Average Loss: 3.9540
Validation Loss: 4.0021
Epoch 8, Batch 100, Loss: 3.7927
Epoch 8, Batch 200, Loss: 3.8502
Epoch 8 completed in 45.71s, Average Loss: 3.8260
Validation Loss: 3.9152
Epoch 9, Batch 100, Loss: 3.8122
Epoch 9, Batch 200, Loss: 3.6185
Epoch 9 completed in 45.74s, Average Loss: 3.7111
Validation Loss: 3.8550
Epoch 10, Batch 100, Loss: 3.5541
Epoch 10, Batch 200, Loss: 3.3463
Epoch 10 completed in 45.65s, Average Loss: 3.6070
```
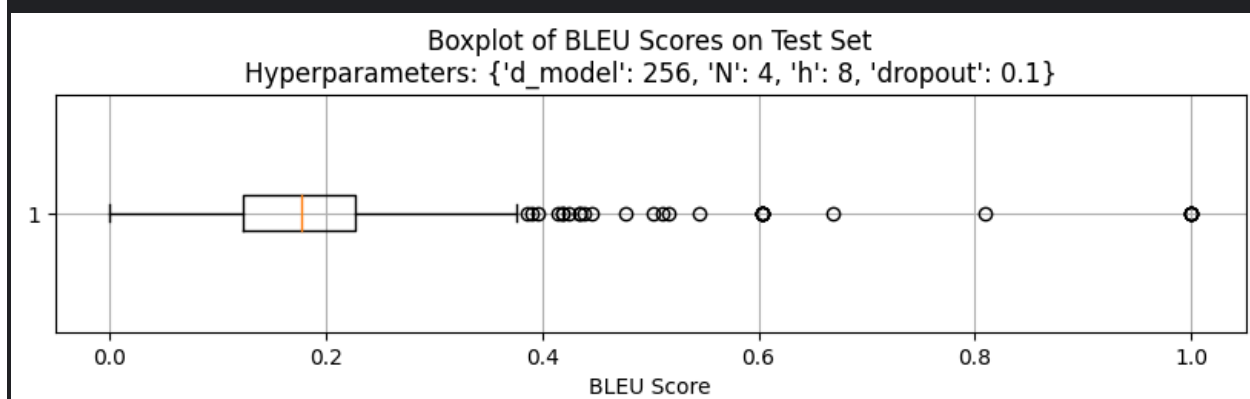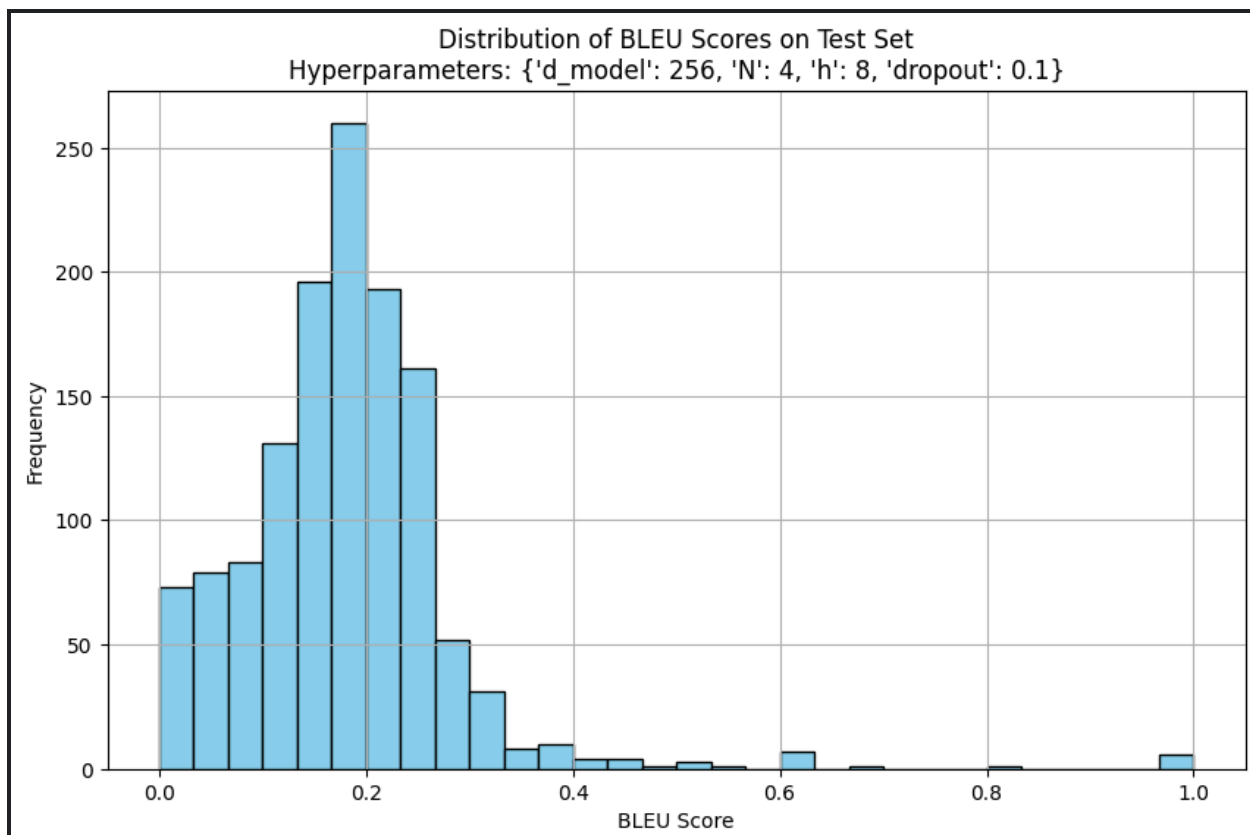
```
Validation Loss: 3.7825
Model saved to transformer_model_0.pt
Average BLEU score: 0.1805
BLEU scores saved to testbleu_0.txt
```
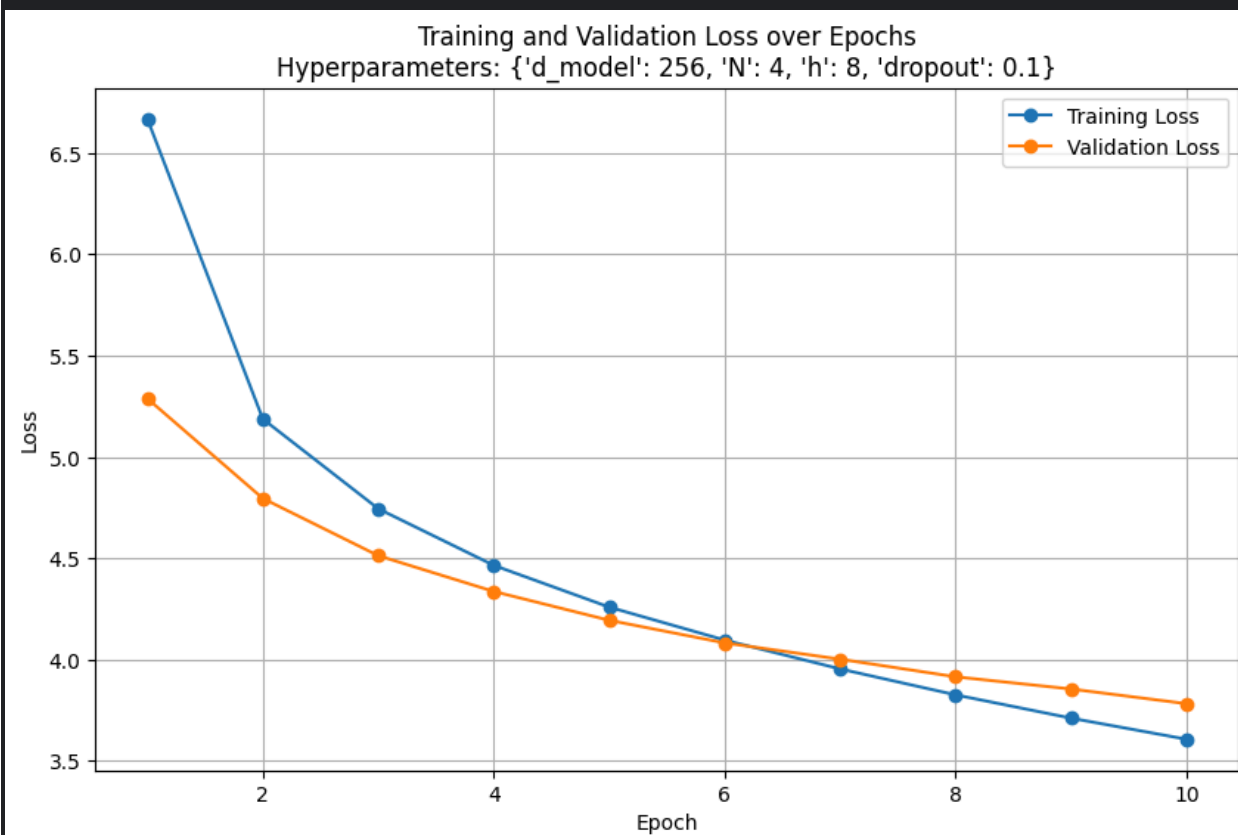
Training and Validation Loss over Epochs
Hyperparameters: {'d_model': 256, 'N': 4, 'h': 8, 'dropout': 0.1}


```
   Prediction  BLEU_Score
0  quand je me suis allé à ma mère , j'ai commenc...   0.178918
1     j'étais un moment dans le <UNK> de la guerre .   0.232523
2                    elle était un <UNK> < /talkid >   0.112457
3  maintenant , le premier dans son premier premi...   0.082145
4  quand je suis entendu , je me suis donc <UNK> ...   0.266259
```

Distribution of BLEU Scores on Test Set
Hyperparameters: {'d_model': 256, 'N': 4, 'h': 8, 'dropout': 0.1}

Boxplot of BLEU Scores on Test Set
Hyperparameters: {'d_model': 256, 'N': 4, 'h': 8, 'dropout': 0.1}


```
BLEU Score Summary Statistics for Hyperparameters: {'d_model': 256, 'N': 4,
'h': 8, 'dropout': 0.1}
count    1305.000000
mean        0.180546
std         0.105971
min         0.000000
25%         0.124274
50%         0.1775
```
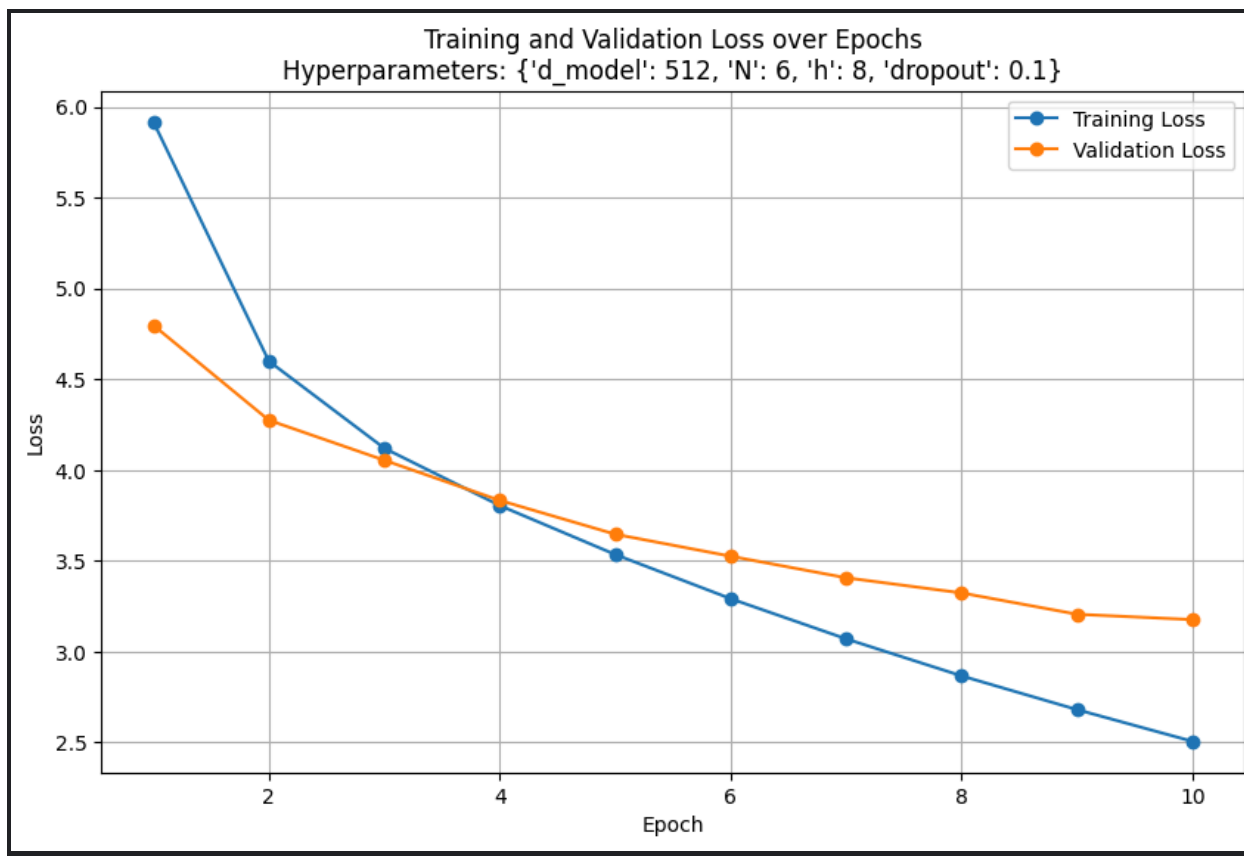
```
Training model with hyperparameters: {'d_model': 512, 'N': 6, 'h': 8,
```

```
'dropout': 0.1}
Epoch 1, Batch 100, Loss: 5.6721
Epoch 1, Batch 200, Loss: 5.1796
Epoch 1 completed in 104.58s, Average Loss: 5.9170
Validation Loss: 4.7962
Epoch 2, Batch 100, Loss: 4.5093
Epoch 2, Batch 200, Loss: 4.2901
Epoch 2 completed in 104.81s, Average Loss: 4.5986
Validation Loss: 4.2733
Epoch 3, Batch 100, Loss: 4.0714
Epoch 3, Batch 200, Loss: 3.8611
Epoch 3 completed in 104.81s, Average Loss: 4.1188
Validation Loss: 4.0527
Epoch 4, Batch 100, Loss: 4.0910
Epoch 4, Batch 200, Loss: 3.8563
Epoch 4 completed in 104.71s, Average Loss: 3.8029
Validation Loss: 3.8314
Epoch 5, Batch 100, Loss: 3.6262
Epoch 5, Batch 200, Loss: 3.4225
Epoch 5 completed in 104.37s, Average Loss: 3.5334
Validation Loss: 3.6450
Epoch 6, Batch 100, Loss: 3.3267
Epoch 6, Batch 200, Loss: 3.1293
Epoch 6 completed in 104.59s, Average Loss: 3.2899
Validation Loss: 3.5239
Epoch 7, Batch 100, Loss: 3.0725
Epoch 7, Batch 200, Loss: 3.0603
Epoch 7 completed in 104.45s, Average Loss: 3.0686
Validation Loss: 3.4051
Epoch 8, Batch 100, Loss: 2.7463
Epoch 8, Batch 200, Loss: 2.9080
Epoch 8 completed in 104.46s, Average Loss: 2.8647
Validation Loss: 3.3217
Epoch 9, Batch 100, Loss: 2.8002
Epoch 9, Batch 200, Loss: 2.6939
Epoch 9 completed in 104.39s, Average Loss: 2.6788
Validation Loss: 3.2043
Epoch 10, Batch 100, Loss: 2.3713
Average BLEU score: 0.2165
BLEU scores saved to testbleu_1.txt
```
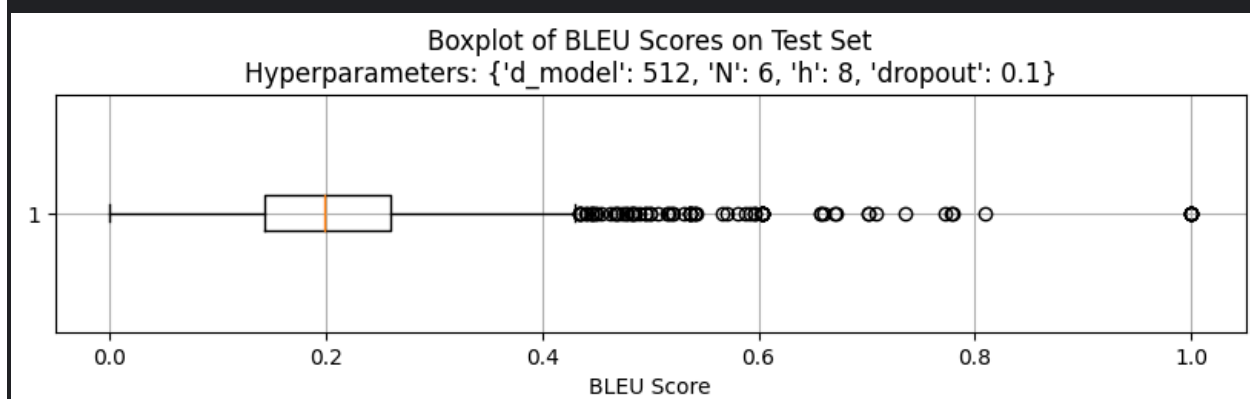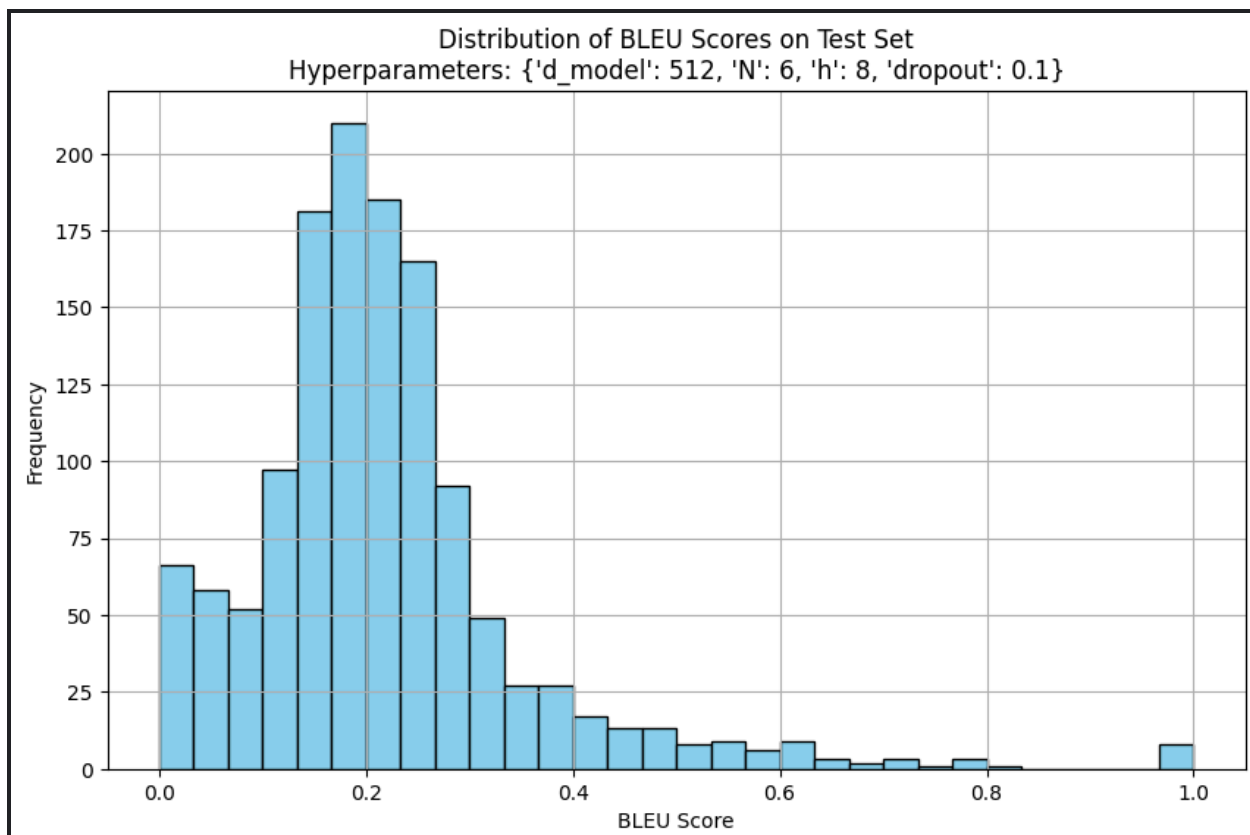
```
Training model with hyperparameters: {'d_model': 512, 'N': 6, 'h': 8,
'dropout': 0.1}
Epoch 1, Batch 100, Loss: 5.6721
```

```
Epoch 1, Batch 200, Loss: 5.1796
Epoch 1 completed in 104.58s, Average Loss: 5.9170
Validation Loss: 4.7962
Epoch 2, Batch 100, Loss: 4.5093
Epoch 2, Batch 200, Loss: 4.2901
Epoch 2 completed in 104.81s, Average Loss: 4.5986
Validation Loss: 4.2733
Epoch 3, Batch 100, Loss: 4.0714
Epoch 3, Batch 200, Loss: 3.8611
Epoch 3 completed in 104.81s, Average Loss: 4.1188
Validation Loss: 4.0527
Epoch 4, Batch 100, Loss: 4.0910
Epoch 4, Batch 200, Loss: 3.8563
Epoch 4 completed in 104.71s, Average Loss: 3.8029
Validation Loss: 3.8314
Epoch 5, Batch 100, Loss: 3.6262
Epoch 5, Batch 200, Loss: 3.4225
Epoch 5 completed in 104.37s, Average Loss: 3.5334
Validation Loss: 3.6450
Epoch 6, Batch 100, Loss: 3.3267
Epoch 6, Batch 200, Loss: 3.1293
Epoch 6 completed in 104.59s, Average Loss: 3.2899
Validation Loss: 3.5239
Epoch 7, Batch 100, Loss: 3.0725
Epoch 7, Batch 200, Loss: 3.0603
Epoch 7 completed in 104.45s, Average Loss: 3.0686
Validation Loss: 3.4051
Epoch 8, Batch 100, Loss: 2.7463
Epoch 8, Batch 200, Loss: 2.9080
Epoch 8 completed in 104.46s, Average Loss: 2.8647
Validation Loss: 3.3217
Epoch 9, Batch 100, Loss: 2.8002
Epoch 9, Batch 200, Loss: 2.6939
Epoch 9 completed in 104.39s, Average Loss: 2.6788
Validation Loss: 3.2043
Epoch 10, Batch 100, Loss: 2.3713
Average BLEU score: 0.2165
BLEU scores saved to testbleu_1.txt
```

Training and Validation Loss over Epochs
Hyperparameters: {'d_model': 512, 'N': 6, 'h': 8, 'dropout': 0.1}

Prediction BLEU_Score

```
0  quand j'étais dans mon mari , j'ai vu mon prem...    0.156812
1   j'étais un <UNK> en psychologie de psychologie .    0.182277
2         elle était une femme <UNK> < /description >    0.149696
3  maintenant , le dr dans son premier ministre d...    0.038785
4  quand j'ai entendu cette question , j'étais si...    0.247746
```

Distribution of BLEU Scores on Test Set
Hyperparameters: {'d_model': 512, 'N': 6, 'h': 8, 'dropout': 0.1}

Boxplot of BLEU Scores on Test Set
Hyperparameters: {'d_model': 512, 'N': 6, 'h': 8, 'dropout': 0.1}

```
BLEU Score Summary Statistics for Hyperparameters: {'d_model': 512, 'N': 6,
'h': 8, 'dropout': 0.1}
count    1305.000000
mean        0.216528
std         0.135405
min         0.000000
25%         0.143317
50%         0.198755
75%         0.259719
max         1.000000
```

```
Training model with hyperparameters: {'d_model': 256, 'N': 4, 'h': 4,
'dropout': 0.2}
Epoch 1, Batch 100, Loss: 6.4750
Epoch 1, Batch 200, Loss: 5.6550
Epoch 1 completed in 43.44s, Average Loss: 6.6516
Validation Loss: 5.3216
Epoch 2, Batch 100, Loss: 5.3275
Epoch 2, Batch 200, Loss: 4.9885
Epoch 2 completed in 43.56s, Average Loss: 5.2608
Validation Loss: 4.8280
Epoch 3, Batch 100, Loss: 4.8978
Epoch 3, Batch 200, Loss: 4.8167
Epoch 3 completed in 43.47s, Average Loss: 4.8205
Validation Loss: 4.5440
Epoch 4, Batch 100, Loss: 4.5035
Epoch 4, Batch 200, Loss: 4.4852
Epoch 4 completed in 43.47s, Average Loss: 4.5452
Validation Loss: 4.3637
Epoch 5, Batch 100, Loss: 4.1785
Epoch 5, Batch 200, Loss: 4.3246
Epoch 5 completed in 43.89s, Average Loss: 4.3467
Validation Loss: 4.2382
Epoch 6, Batch 100, Loss: 4.1981
Epoch 6, Batch 200, Loss: 4.1703
Epoch 6 completed in 43.85s, Average Loss: 4.1888
Validation Loss: 4.1297
Epoch 7, Batch 100, Loss: 4.1400
Epoch 7, Batch 200, Loss: 4.1160
Epoch 7 completed in 43.48s, Average Loss: 4.0518
Validation Loss: 4.0421
Epoch 8, Batch 100, Loss: 3.8872
Epoch 8, Batch 200, Loss: 3.7491
Epoch 8 completed in 43.47s, Average Loss: 3.9313
Validation Loss: 3.9554
Epoch 9, Batch 100, Loss: 3.8363
Epoch 9, Batch 200, Loss: 3.7263
Epoch 9 completed in 43.40s, Average Loss: 3.8202
Validation Loss: 3.8924
Epoch 10, Batch 100, Loss: 3.8014
Epoch 10, Batch 200, Loss: 3.6610
Epoch 10 completed in 43.43s, Average Loss: 3.7185
Validation Loss: 3.8143
Model saved to transformer_model_2.pt
Average BLEU score: 0.1787
BLEU scores saved to testbleu_2.txt
```
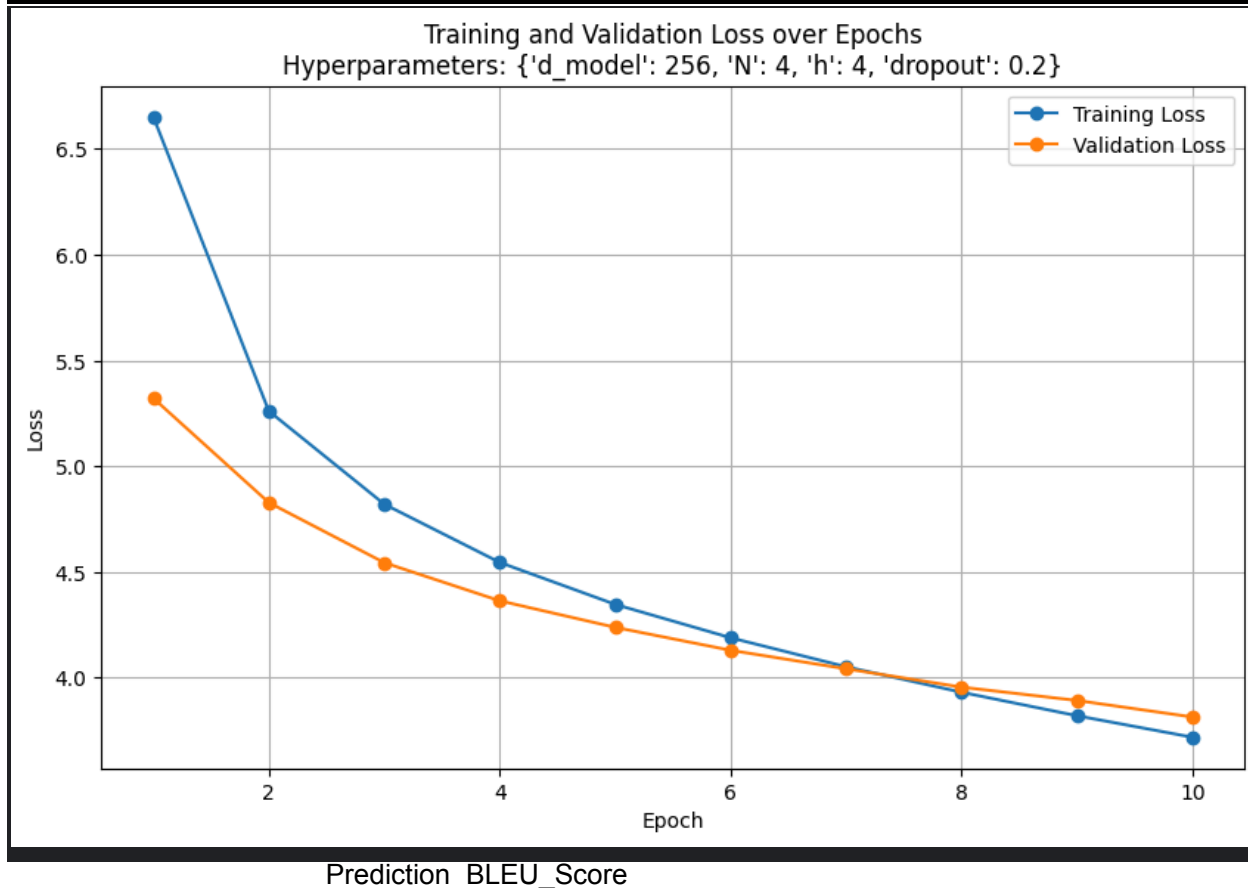
```
Training model with hyperparameters: {'d_model': 256, 'N': 4, 'h': 4,
'dropout': 0.2}
Epoch 1, Batch 100, Loss: 6.4750
Epoch 1, Batch 200, Loss: 5.6550
Epoch 1 completed in 43.44s, Average Loss: 6.6516
Validation Loss: 5.3216
Epoch 2, Batch 100, Loss: 5.3275
Epoch 2, Batch 200, Loss: 4.9885
Epoch 2 completed in 43.56s, Average Loss: 5.2608
Validation Loss: 4.8280
Epoch 3, Batch 100, Loss: 4.8978
Epoch 3, Batch 200, Loss: 4.8167
Epoch 3 completed in 43.47s, Average Loss: 4.8205
Validation Loss: 4.5440
Epoch 4, Batch 100, Loss: 4.5035
Epoch 4, Batch 200, Loss: 4.4852
Epoch 4 completed in 43.47s, Average Loss: 4.5452
Validation Loss: 4.3637
Epoch 5, Batch 100, Loss: 4.1785
Epoch 5, Batch 200, Loss: 4.3246
Epoch 5 completed in 43.89s, Average Loss: 4.3467
Validation Loss: 4.2382
Epoch 6, Batch 100, Loss: 4.1981
Epoch 6, Batch 200, Loss: 4.1703
Epoch 6 completed in 43.85s, Average Loss: 4.1888
Validation Loss: 4.1297
Epoch 7, Batch 100, Loss: 4.1400
Epoch 7, Batch 200, Loss: 4.1160
Epoch 7 completed in 43.48s, Average Loss: 4.0518
Validation Loss: 4.0421
Epoch 8, Batch 100, Loss: 3.8872
Epoch 8, Batch 200, Loss: 3.7491
Epoch 8 completed in 43.47s, Average Loss: 3.9313
Validation Loss: 3.9554
Epoch 9, Batch 100, Loss: 3.8363
Epoch 9, Batch 200, Loss: 3.7263
Epoch 9 completed in 43.40s, Average Loss: 3.8202
Validation Loss: 3.8924
Epoch 10, Batch 100, Loss: 3.8014
Epoch 10, Batch 200, Loss: 3.6610
Epoch 10 completed in 43.43s, Average Loss: 3.7185
Validation Loss: 3.8143
Model saved to transformer_model_2.pt
Average BLEU score: 0.1787
```
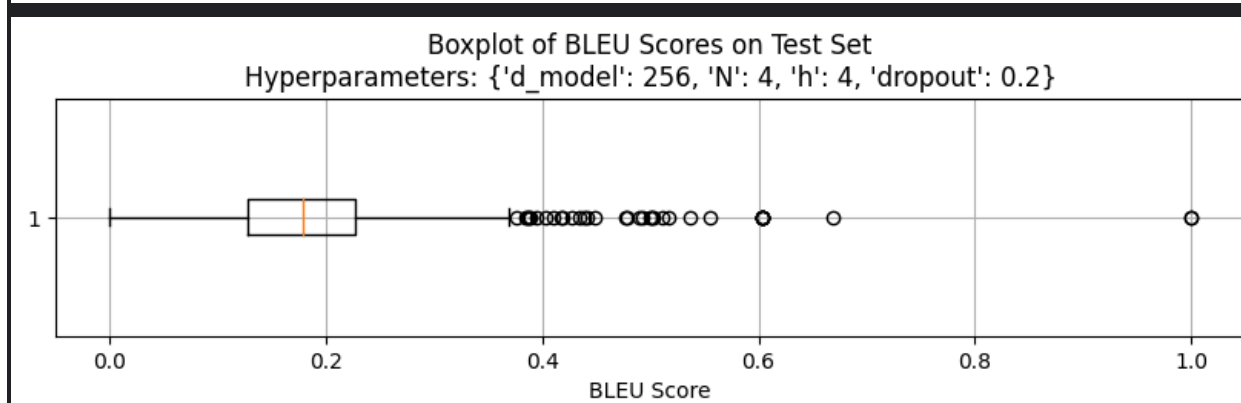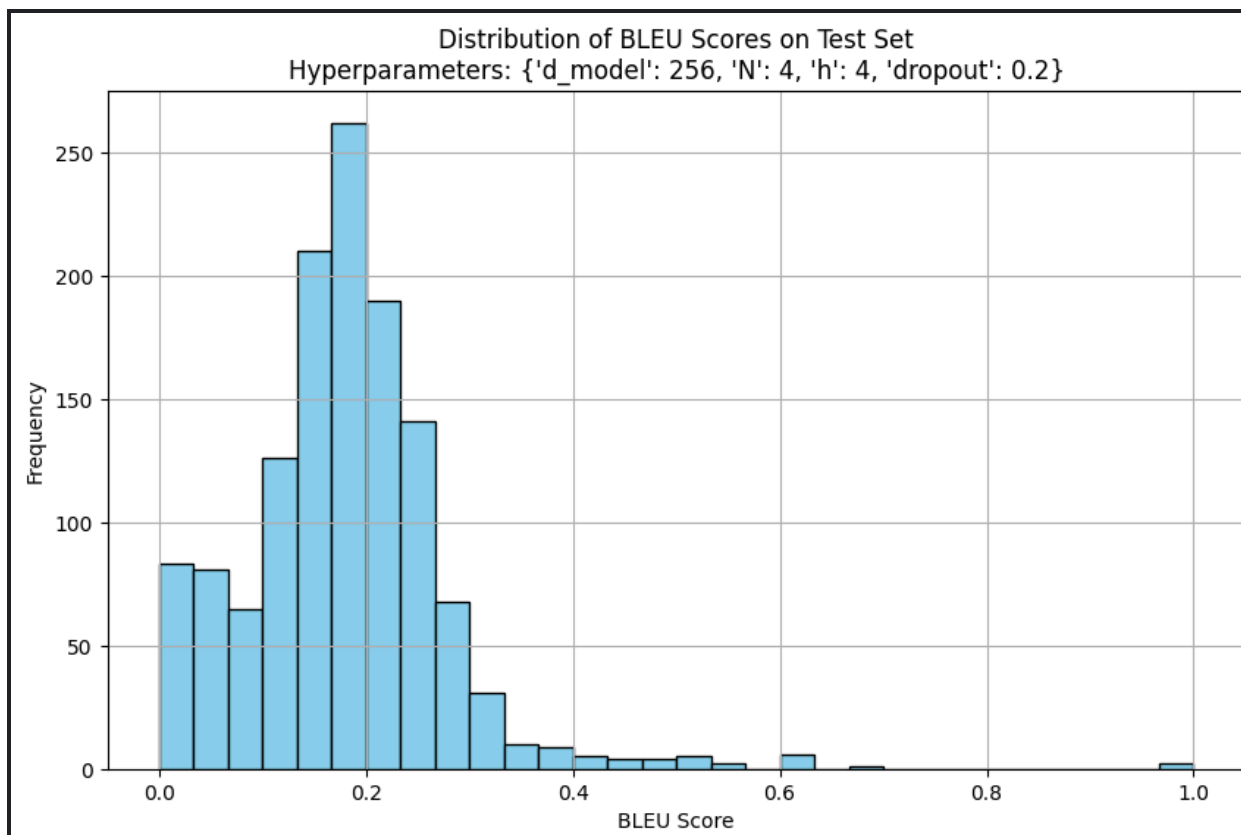
Training and Validation Loss over Epochs
Hyperparameters: {'d_model': 256, 'N': 4, 'h': 4, 'dropout': 0.2}

Prediction BLEU_Score

```
0  quand j'étais dans mon père , j'ai vu mon prem...    0.193324
1     j'étais une photo de la maison dans les pays .    0.232523
2                       elle était un <UNK> < /url .    0.133734
3  maintenant , les <UNK> dans sa femme et <UNK> ...    0.029215
4  quand je suis donc , je suis donc je suis donc...    0.189681
```

Distribution of BLEU Scores on Test Set
Hyperparameters: {'d_model': 256, 'N': 4, 'h': 4, 'dropout': 0.2}

Boxplot of BLEU Scores on Test Set
Hyperparameters: {'d_model': 256, 'N': 4, 'h': 4, 'dropout': 0.2}


```
BLEU Score Summary Statistics for Hyperparameters: {'d_model': 256, 'N': 4,
'h': 4, 'dropout': 0.2}
count    1305.000000
mean        0.178715
std         0.097108
min         0.000000
25%         0.128343
50%         0.178918
75%         0.227648
max         1.000000
Name: BLEU_Score, dtype: float64
```

```
Training model with hyperparameters: {'d_model': 128, 'N': 2, 'h': 4,
'dropout': 0.1}
Epoch 1, Batch 100, Loss: 7.8482
Epoch 1, Batch 200, Loss: 6.8044
Epoch 1 completed in 27.56s, Average Loss: 7.7716
Validation Loss: 6.1768
Epoch 2, Batch 100, Loss: 6.0990
Epoch 2, Batch 200, Loss: 5.9329
Epoch 2 completed in 27.67s, Average Loss: 6.0328
Validation Loss: 5.4717
Epoch 3, Batch 100, Loss: 5.5549
Epoch 3, Batch 200, Loss: 5.3932
Epoch 3 completed in 27.73s, Average Loss: 5.5445
Validation Loss: 5.1570
Epoch 4, Batch 100, Loss: 5.2491
Epoch 4, Batch 200, Loss: 5.2732
Epoch 4 completed in 27.95s, Average Loss: 5.2409
Validation Loss: 4.9512
Epoch 5, Batch 100, Loss: 5.0403
Epoch 5, Batch 200, Loss: 4.8846
Epoch 5 completed in 27.93s, Average Loss: 5.0236
Validation Loss: 4.8060
Epoch 6, Batch 100, Loss: 4.9417
Epoch 6, Batch 200, Loss: 4.7996
Epoch 6 completed in 28.10s, Average Loss: 4.8527
Validation Loss: 4.6763
Epoch 7, Batch 100, Loss: 4.6878
Epoch 7, Batch 200, Loss: 4.5556
Epoch 7 completed in 27.99s, Average Loss: 4.7138
Validation Loss: 4.5821
Epoch 8, Batch 100, Loss: 4.4442
Epoch 8, Batch 200, Loss: 4.4640
Epoch 8 completed in 28.10s, Average Loss: 4.5969
Validation Loss: 4.4959
Epoch 9, Batch 100, Loss: 4.5629
Epoch 9, Batch 200, Loss: 4.4726
Epoch 9 completed in 27.96s, Average Loss: 4.4958
Validation Loss: 4.4231
Epoch 10, Batch 100, Loss: 4.4913
Epoch 10, Batch 200, Loss: 4.4313
Epoch 10 completed in 27.63s, Average Loss: 4.4081
Validation Loss: 4.3590
Model saved to transformer_model_3.pt
Average BLEU score: 0.1647
```
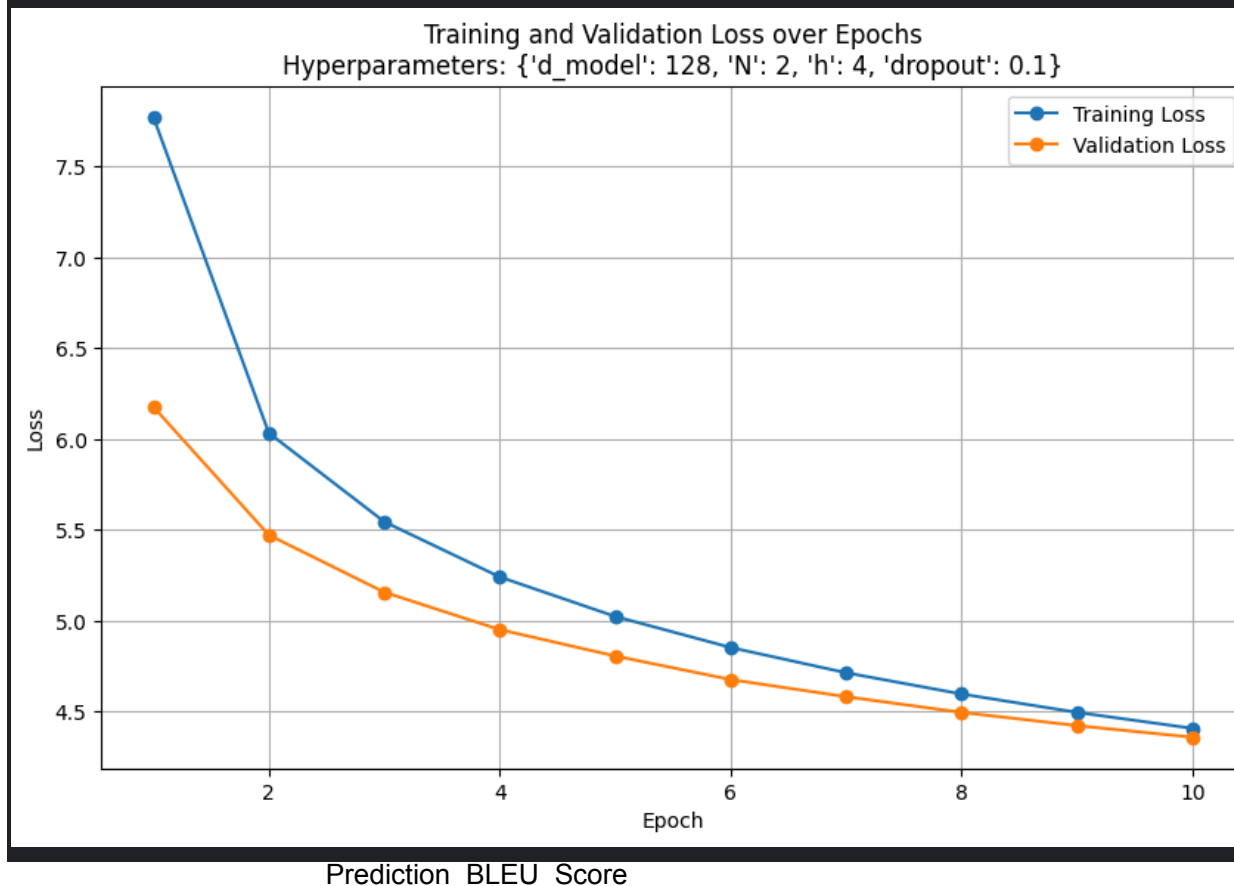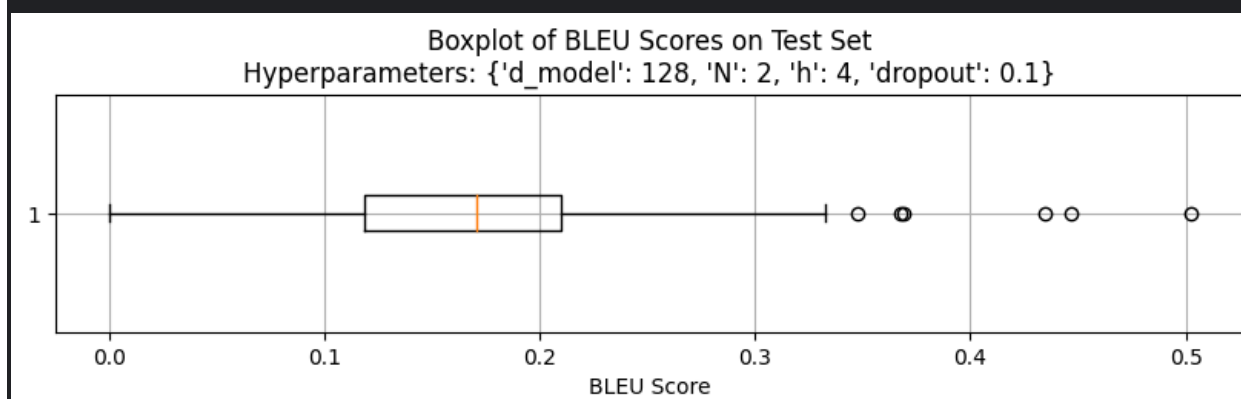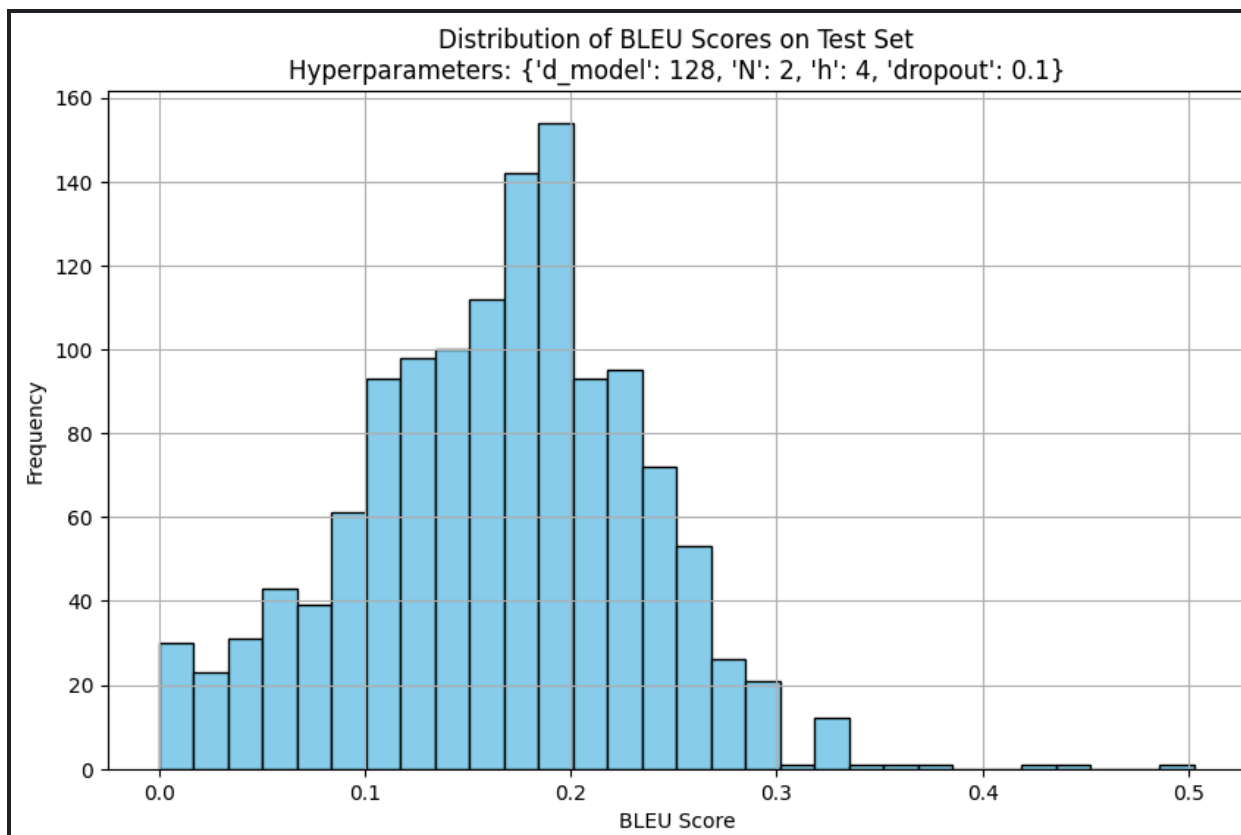
Training and Validation Loss over Epochs
Hyperparameters: {'d_model': 128, 'N': 2, 'h': 4, 'dropout': 0.1}

Prediction BLEU_Score

	Prediction	BLEU_Score
0	quand je suis allé dans mon père , je suis <UN...	0.249635
1	je suis une fois dans un peu de la première .	0.194068
2	elle a un <UNK> < /talkid >	0.112457
3	maintenant , les <UNK> de la <UNK> et <UNK> <U...	0.021878
4	quand je suis donc , je suis <UNK> < /url >	0.230787

Distribution of BLEU Scores on Test Set
Hyperparameters: {'d_model': 128, 'N': 2, 'h': 4, 'dropout': 0.1}

Boxplot of BLEU Scores on Test Set
Hyperparameters: {'d_model': 128, 'N': 2, 'h': 4, 'dropout': 0.1}


```
BLEU Score Summary Statistics for Hyperparameters: {'d_model': 128, 'N': 2,
'h': 4, 'dropout': 0.1}
count    1305.000000
mean        0.164696
std         0.069352
min         0.000000
25%         0.118805
50%         0.170367
75%         0.209576
max         0.502659
Name: BLEU_Score, dtype: float64
```

```
Results for Hyperparameters: {'d_model': 256, 'N': 4, 'h': 8, 'dropout': 0.1}
Average Training Loss: 3.6070
Average Validation Loss: 3.7825
Average BLEU Score: 0.1805

Results for Hyperparameters: {'d_model': 512, 'N': 6, 'h': 8, 'dropout': 0.1}
Average Training Loss: 2.5040
Average Validation Loss: 3.1747
Average BLEU Score: 0.2165

Results for Hyperparameters: {'d_model': 256, 'N': 4, 'h': 4, 'dropout': 0.2}
Average Training Loss: 3.7185
Average Validation Loss: 3.8143
Average BLEU Score: 0.1787

Results for Hyperparameters: {'d_model': 128, 'N': 2, 'h': 4, 'dropout': 0.1}
Average Training Loss: 4.4081
Average Validation Loss: 4.3590
Average BLEU Score: 0.1647
```

The bleu score is observed between 0.11 to 0.299
By varying different parameters.

The low bleu score may be due to limited compute,
architecture and data that is used.