Model Report (Metrics Summary of best model)

Based on the training and evaluation of the BERT-based sentiment classifier, summary of the key metrics:

BERT-based Sentiment Classifier Performance (on Test Set after 4 Epochs):

Metric	Negative (0)	Neutral (1)	Positive (2)	Macro Avg	Weighted Avg
Precision	0.99	0.99	0.97	0.98	0.98
Recall	0.97	0.99	0.99	0.98	0.98
F1-Score	0.98	0.99	0.98	0.98	0.98
Support	214	437	266	917	917

Overall Accuracy: 0.98

Validation AUC-ROC (One-vs-Rest): 0.999

Interpretation:

- **High Accuracy:** The model achieved a high overall accuracy of 98% on the test set, indicating that it correctly classified a large majority of the reviews.

- **Strong Precision and Recall:** The high precision and recall scores across all three classes (Negative, Neutral, Positive) suggest that the model is effective at identifying instances of each sentiment and minimizing both false positives and false negatives.

- **High F1-Score:** The F1-scores, which are the harmonic mean of precision and recall, are also very high, indicating a good balance between precision and recall for each class.

- **Excellent AUC-ROC:** The AUC-ROC score of 0.999 is very close to 1, which signifies that the model has an excellent ability to distinguish between the different sentiment classes.

Comparison with FastText-based Models:

The BERT-based model significantly outperformed the FastText-based LSTM, GRU, and BiLSTM models, which had validation accuracies and F1-scores in the range of 0.91-0.92. This highlights the superior capability of pre-trained transformer models like BERT for capturing complex semantic relationships in text, leading to more accurate sentiment classification.

Conclusion:

The fine-tuned BERT sentiment classifier demonstrates strong performance on this dataset, achieving high accuracy and robust F1-scores across all sentiment classes. The high AUC-ROC further confirms its excellent discriminatory power. This model appears to be well-suited for the task of classifying the sentiment of HP printer reviews.