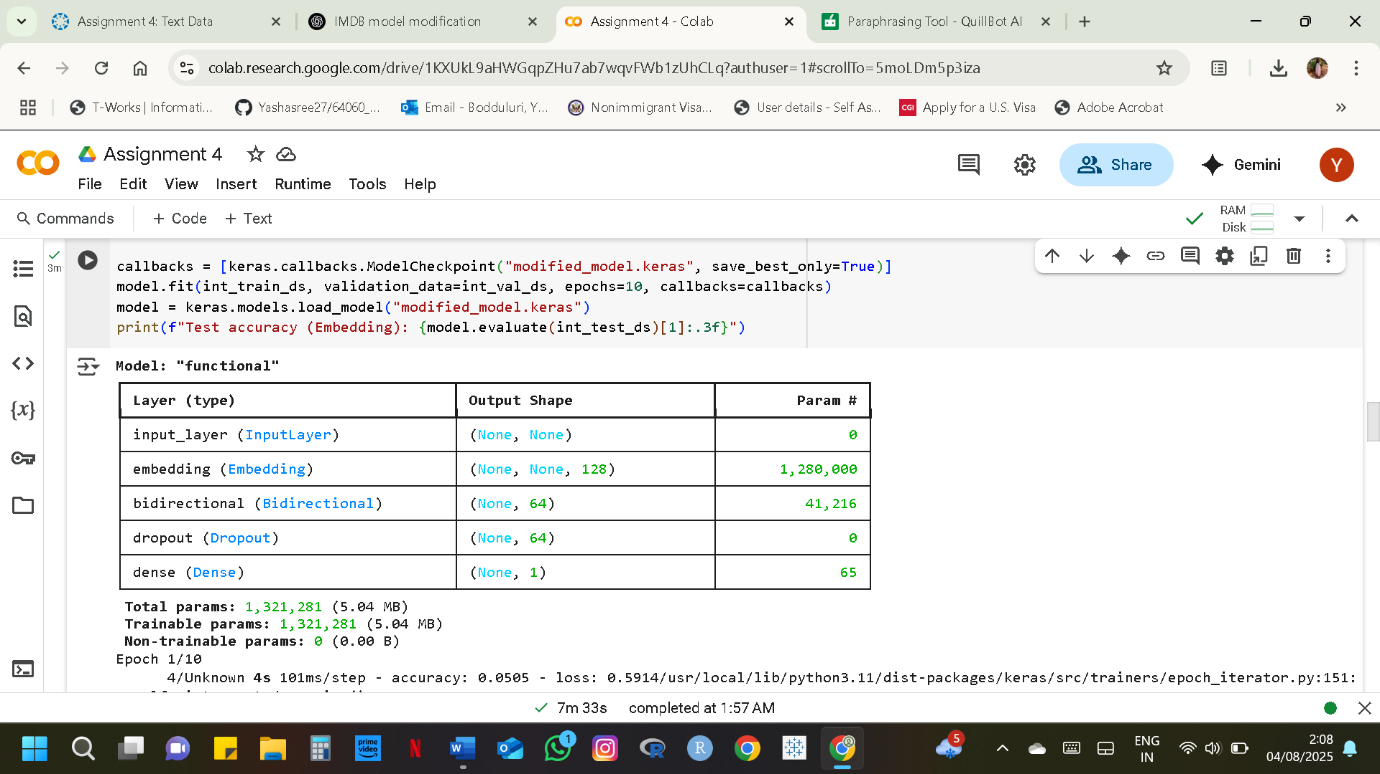
**Assignment 4**

Using the IMDB movie review dataset, I looked into the performance of two separate word embedding techniques for sentiment classification, especially in a training environment with limitations.   
I limited the training set to just 100 reviews, capped each review at 150 tokens, and only used the top 10,000 most frequently occurring words in the vocabulary in order to evaluate how well models can generalize from limited data. The standard 25,000-sample IMDB test set was used for testing, and 10,000 samples were used to create the validation set.   
A bidirectional LSTM layer, dropout, and a final dense output layer were all part of the architecture for both models.

A common trainable embedding layer that was initially created from scratch was used in the first model.

Considering the extremely small amount of training data, this model's test accuracy of 81.2% is significant.

The second model used 100-dimensional vectors of pretrained GloVe word embeddings and maintained these embeddings constant throughout training. With a test accuracy of 82.3%, this model slightly outperformed the trainable embedding model.

**IMDB Dataset**

An overview of IMDB Dataset is shown in the image. An Input Layer, which supports input sequences of variable length, was used first.

Each input token is then transformed into a 128-dimensional dense vector by the `Embedding` layer, resulting 1,280,000 trainable parameters.

A Bidirectional layer comes next, which contributes 41,216 parameters and wraps a recurrent layer (such as LSTM or GRU) to generate a 64-dimensional vector by combining forward and backward passes.

After that, a Dropout layer randomly drops units during training to help avoid overfitting.  
Lastly, a Dense layer with 65 trainable parameters produces a single value appropriate for binary classification. There are 1,321,281 trainable parameters in the model overall, and none of them are non-trainable.

**A screenshot of a computer

AI-generated content may be incorrect.For GloVe dataset**

The summary of a GloVe Dataset is shown in this image. It begins with an input layer that can handle input sequences of varying length.

Each token has been assigned to a 100-dimensional vector by the embedding layer, which has a total of 1,000,000 non-trainable parameters. This means that the embedding weights, such as GloVe or Word2Vec, are pre-trained and frozen during training.

After that, a bidirectional layer contributes 34,048 trainable parameters and produces a 64-dimensional representation after processing the sequence both forward and backward.

These results indicate how well pretrained embeddings work for transfer learning, especially in settings with limited resources. The relationships captured in the pretrained vectors gave the GloVe model a slight but significant generalization advantage, despite the fact that both models performed well.

Therefore, we conclude that when training data is limited, the use of pretrained embeddings such as GloVe can significantly improve model performance with minimal change.