DISTILLBERT MODEL FOR UTTERANCE CLASSIFICATION

# Model Training

The Utterances were classified using a DistillBert pretrained model from the Python Transformers library. The Transformers pretrained tokenizer is used to tokenize utterances.

Before passing through the model, several data cleaning processes were applied to the utterances as follows:

### Data Cleaning steps -

1. Lowercase the text - For uniformity and accuracy, the text of the utterances is lowercased.
2. Lemmatization - Lemmatization is the process of transforming the text of an utterance into a lemma.
3. Correction of spelling errors - The Python spell correction library is used to fix spelling errors in utterances.

### Model, Loss functions and matrix details -

1. The model is trained using the Python Transformers model *TFdistillbertforSequenceClassification*.
2. During model training, the Sparse Categorical Cross entropy Loss is utilised as a loss function.
3. Adam Optimizer with a learning rate of 0.00002 is utilized as a metric. The batch size is 32 and the number of epochs is kept at 30.

# Model Testing

To test the model, all of the test data from the seven domains is pooled into a single file, which is then run through the model to determine the combined accuracy. In addition, all of the individual accuracies have been calculated. During validation, the accuracy score is used as a metric.

# Deployment

The Python flask library has been used to deploy the trained model on the server. The web app is set up so that if an utterance is passed, the app will return the intent as well as the top three confidence scores.