**Report 1:**  
Shilo Avital 206487407 Tamar Michelson 323805861

**1. Dataset**

The model was trained on 1000 sequences: 500 positive and 500 negative examples. The same number of samples were used for the test set.

**2. Model Architecture**

The model used is an LSTM-based acceptor with an embedding layer, followed by an LSTMCell and a multi-layer perceptron with one hidden layer (16 units, ReLU activation) and a final output layer for binary classification.

**3. Training Details**

• Embedding dimension: 16  
• Hidden dimension: 128  
• Optimizer: Adam  
• Loss: Binary Cross-Entropy with Logits  
• Epochs: 10  
• Batch size: 64  
• Learning rate: 0.001

**4. Performance**

The model was evaluated on a separate test set. The final test accuracy was approximately reported per epoch and converged above 0.9. Training time and step count are printed to stdout.

**5. Observations and Fixes**

Some key aspects in model development included proper masking of padded elements in sequences to prevent state updates for padding. Initial trials showed high training accuracy but required adjustments for generalization such as tuning learning rate and embedding dimension.

