IMPLEMENTATION DOCUMENTATION

1. **CHALLENGES**
   - Dataset size issues:
     a. The dataset provided in the github was too large and it would have taken a lot of time for it to download
     b. **Solution:** Selected another open source dataset of .gz format
   - **Dataset extraction issues:**
     a. The `tar.extractall()` method was initially facing permission errors.
     b. **Solution:** Used `tempfile.mkdtemp()` to create a dedicated extraction directory.
   - **Labeling ambiguity:**
     a) File names did not always contain clear labels (`spoof` or `fake`).
     b) **Solution:** Used case-insensitive checks for keywords and ensured balanced dataset distribution.
   - **Model convergence issues:**
     a) Training loss sometimes stagnated.
     b) **Solution:** Tuned learning rate and implemented `EarlyStopping`.

2. **ASSUMPTIONS MADE**
   - All the files are either labelled bonafide or spoofed
   - All files are well labelled
   - Augmentation improves model robustness without introducing artifacts

3. **MODEL SELECTION**
   - Convolutional Recurrent Neural Network (CRNN) with Attention was chosen for its ability to capture both spatial and temporal patterns in audio spectrograms.
   - Bidirectional GRU (BGRU) layers enhance sequential feature learning.
   - Multi-Head Attention improves feature representation by focusing on key aspects of the spectrogram.

4. **FUTURE IMPROVEMENTS**
   - Load a dataset that is properly labelled.
   - Introduce noise so that the model can self learn to remove the noisy pattern.
   - Improve performance in various types of audios.

- Combine CRNN with Transformer-based architectures for enhanced feature extraction.

5. **REFLECTION QUESTIONS**
   - Handling the dataset was indeed a challenge as segregating so many unlabeled data was a task itself which failed.
   - The combined method of self attention with CRNN could generate better results compared to the case studies as it performs well when noise is introduced.
   - For deployment prospects, the model has to train continuously so that it can keep updated of the new variations of sound.
   - Need for synthetic data augmentation to enhance robustness against unseen attack types.