

## Model & Dataset Selection

### Selected Models for Forgery Detection

Based on the requirements (AI-generated speech detection, real-time potential, real conversation analysis), I've chosen these three models:

#### 1. RawNet2

- Key Innovation: Uses raw waveform input instead of spectrograms, leveraging deep residual networks for feature extraction.
- Performance: Strong performance on ASVspoof datasets with high EER (Equal Error Rate) reduction.
- Why It's Promising: Efficient feature learning directly from waveforms, reducing preprocessing overhead.
- Limitations: Requires large-scale training data for generalization.

#### 2. LCNN (Light Convolutional Neural Network)

- Key Innovation: Uses depth-wise separable convolutions with max-feature mapping for robustness.
- Performance: Achieves high accuracy in detecting synthetic speech in benchmark datasets.
- Why It's Promising: Lightweight and computationally efficient, making it suitable for real-time applications.
- Limitations: Performance can vary across unseen deepfake generation methods.

#### 3. AASIST (Audio Anti-Spoofing using Integrated Spectro-Temporal Features)

- Key Innovation: Jointly models temporal and spectral information using convolutional recurrent networks.
- Performance: State-of-the-art accuracy on ASVspoof2021 dataset.
- Why It's Promising: Designed explicitly for anti-spoofing, incorporating frequency and time-based detection.
- Limitations: Higher computational cost compared to LCNN.

### Chosen Model for Implementation

I'll go with RawNet2, as it eliminates the need for handcrafted spectrogram features and operates directly on raw audio, making it suitable for diverse speech scenarios.