**PART 1: RESEARCH & SELECTION**

Here are **three promising models/approaches** for detecting AI-generated human speech, considering **real-time detection, conversational analysis, and effectiveness**:

**Wav2Vec 2.0-Based Deepfake Detection**

🔹 **Why?**

* Self-supervised learning on speech audio makes it **effective for deepfake detection**.
* Pretrained on large datasets → **Robust against AI-generated voices**.
* Can be **fine-tuned for real-time detection** with lightweight models.

🔹 **How It Works?**

* Converts audio into latent representations.
* Uses a Transformer-based model to classify human vs. AI-generated speech.
* Fine-tuning with deepfake datasets improves **accuracy**.

🔹 **Potential for Real-Time Detection:** ✅  
🔹 **Strength in Real Conversations:** ✅

**SPECTROGRAM + CNN-Based Approach**

🔹 **Why?**

* AI-generated speech has **subtle spectral anomalies** visible in spectrograms.
* **Convolutional Neural Networks (CNNs)** are excellent for image-based classification.

🔹 **How It Works?**

1. Convert audio signals into **Mel-Spectrograms**.
2. Use a **CNN model** (ResNet, EfficientNet, etc.) to classify real vs. deepfake.
3. Fine-tune with deepfake datasets (e.g., ASVspoof, FakeAVCeleb).

🔹 **Potential for Real-Time Detection:** ⚠️ (Depends on optimization)  
🔹 **Strength in Real Conversations:** ✅

**Raw Audio-Based LSTM/GRU Model**

🔹 **Why?**

* **LSTM/GRU models** excel in detecting **temporal anomalies** in speech.
* AI-generated voices lack **natural variations** in pitch, rhythm, and pauses.

🔹 **How It Works?**

1. Process **raw waveform input** instead of spectrograms.
2. Use **Recurrent Neural Networks (RNNs) with LSTM/GRU** for sequential analysis.
3. Detect **synthetic speech patterns** based on unnatural temporal features.

🔹 **Potential for Real-Time Detection:** ✅ (With optimized inference)  
🔹 **Strength in Real Conversations:** ✅

**Comparison**

| **Model** | **Real-Time Potential** | **Effective in Real Conversations** | **Best For** |
| --- | --- | --- | --- |
| **Wav2Vec 2.0** | ✅ **Yes** (Optimized fine-tuning) | ✅ **Yes** (Pretrained on real speech) | General-purpose deepfake detection |
| **Spectrogram + CNN** | ⚠️ **Possible** (Needs GPU optimization) | ✅ **Yes** (Captures spectral artifacts) | Visual deepfake speech detection |
| **LSTM/GRU on Raw Audio** | ✅ **Yes** (Fast inference) | ✅ **Yes** (Captures temporal anomalies) | Natural speech anomaly detection |