**ALFRED CHINEDU OKORONKWO**
**Title:** *Voice Synthesis: Generating Human-Like Speech from Text*

## Abstract

Recent progress in voice cloning and text-to-speech (TTS) has enabled increasingly natural synthetic speech, yet accurately capturing a target speaker's vocal identity—such as prosody, tone, and timbre—remains challenging. Additionally, TTS-generated audio often contains noise and artifacts that reduce perceptual quality and speaker similarity. This project presents an end-to-end voice cloning and speech enhancement pipeline that combines a pre-trained multilingual TTS model with post-processing denoising techniques. By integrating speaker cloning with audio enhancement, the system produces clearer, more natural synthetic speech that more closely resembles the target speaker.

## System Overview

The pipeline consists of the following stages:

1. **Environment Setup:** Installation of required libraries and dependencies.
2. **Model and Data Preparation:** Downloading a pre-trained multilingual TTS model and a reference speaker WAV file.
3. **Voice Cloning and TTS:** Cloning the speaker's voice and generating speech from input text.
4. **Audio Denoising:** Applying post-processing techniques to reduce noise and artifacts in the generated audio.
5. **Evaluation:** Measuring improvements using acoustic features and qualitative human listening comparisons.

## Dataset and Model

- **Dataset:** LJ001-0001 speaker WAV file from the **LJ Speech Dataset**.
- **Pre-trained Model:** tts_models/multilingual/multi-dataset/your_tts from **Coqui-AI TTS Models**.

## Results

The denoising and enhancement pipeline improved certain acoustic features, notably MFCC coefficient 3, although degradation was observed in MFCC coefficient 13. Despite this trade-off, subjective listening tests indicated that the enhanced audio sounded more natural and perceptually closer to the target speaker.

## Impact and Applications

The proposed system improves the fidelity of TTS-generated speech, producing audio that closely matches the target speaker's vocal characteristics. Potential applications include personalized document reading, assistive technologies for accessibility, audiobook and podcast narration, voice preservation and restoration, human–computer interaction systems, and multilingual content generation while maintaining consistent speaker identity.