Dharshan S
(311521104011)

Final Project
GEN AI

PROJECT TITLE

Conversion of Text to Speech

AGENDA

- PROBLEM STATEMENT
- PROJECT OVERVIEW
- WHO ARE THE END USERS?
- YOUR SOLUTION AND ITS VALUE PROPOSITION
- THE WOW IN YOUR SOLUTION
- MODELLING
- RESULTS

PROBLEM STATEMENT

The project addresses the need for accurate and natural-sounding text-to-speech conversion. It aims to ensure that synthesized speech reflects the original text's meaning and nuances while evaluating and improving the conversion process.

PROJECT OVERVIEW

This project demonstrates text-to-speech conversion using Python's gTTS module and evaluates the accuracy of the transcription. It leverages generative AI to convert written text into natural-sounding speech. The process involves importing libraries, performing conversion, saving audio, playback, and assessing accuracy through character-level comparison. By integrating generative AI techniques, the project showcases the potential for more immersive and realistic speech synthesis.

WHO ARE THE END USERS?

Visually Impaired Individuals: People who are blind or have low vision can use text-to-speech technology to access written content, such as books, websites, and documents, through synthesized speech output.

Language Learners: Individuals learning a new language can utilize text-to-speech systems to improve pronunciation, language comprehension, and listening skills by listening to synthesized speech representations of written text.

Automated Customer Service Systems: Companies deploying automated customer service systems can employ text-to-speech technology to provide voice-based interactions and assistance to customers, enhancing user experience and efficiency.

Educational Institutions: Teachers and students in educational institutions can leverage text-to-speech technology for a variety of purposes, including providing audio versions of textbooks, creating multimedia presentations, and accommodating students with reading disabilities.

Smart Device Users: Consumers using smart devices such as smartphones, tablets, and smart speakers can interact with text-to-speech systems for tasks such as voice search, voice commands, and receiving spoken notifications and reminders.

YOUR SOLUTION AND ITS VALUE PROPOSITION

Solution Overview:

- Text-to-Speech Conversion System with Generative AI Integration

Value Proposition:

- **Enhanced Accessibility:** Our solution empowers visually impaired individuals to access written content through synthesized speech, promoting inclusivity and accessibility in digital communication.
- **Natural and Expressive Speech:** Leveraging generative AI techniques, our system delivers natural-sounding and expressive speech outputs, enhancing user engagement and comprehension.
- **Time Efficiency:** Save time and effort by automating the text-to-speech conversion process, enabling users to quickly convert written text into speech for various applications without manual intervention.
- **Customization Options:** Tailor synthesized speech to individual preferences with customizable parameters such as pitch, speed, and tone, providing users with personalized speech outputs that suit their preferences and needs.
- **Versatility:** Our system's scalability and compatibility with different platforms and devices ensure its versatility for diverse applications, from language learning tools to automated voice interfaces in smart devices.

THE WOW IN YOUR SOLUTION

Natural Human-like Speech: Experience synthesized speech outputs that closely resemble natural human speech, creating a truly immersive and engaging listening experience.
Effortless Customization: Effortlessly customize speech synthesis parameters such as pitch, speed, and tone to create personalized speech outputs tailored to individual preferences and application requirements.
Seamless Integration: Seamlessly integrate the text-to-speech conversion system into existing applications, websites, and devices, enhancing user experience and accessibility without disrupting workflow.
Real-time Feedback: Receive real-time feedback on the accuracy and quality of synthesized speech outputs, enabling users to monitor and adjust speech synthesis parameters for optimal results.
Enhanced Accessibility: Empower individuals with visual impairments or reading difficulties to access written content through synthesized speech, promoting inclusivity and accessibility in digital communication.

MODELLING

Architecture:

The text-to-speech system combines the gTTS module with generative AI techniques. The gTTS module converts text to speech, while the AI component improves speech quality and naturalness.

Training Process:

The generative AI model learns from large human speech datasets, adjusting parameters iteratively to minimize differences between synthesized and real speech.

Loss Functions:

Various loss functions are used to optimize the generative AI model's performance. Common ones include mean squared error (MSE) for speech waveform synthesis and categorical cross-entropy for linguistic feature prediction

Evaluation Metrics:

The system assesses synthesized speech with metrics like word error rate (WER) and naturalness ratings, ensuring accuracy and user satisfaction.

Integration:

The gTTS module handles basic text-to-speech conversion, while the generative AI model enhances speech quality, seamlessly integrated into the conversion pipeline.

RESULTS

The text-to-speech conversion system successfully synthesized speech outputs with high accuracy and naturalness, enhancing user experience and accessibility.

Discriminator Loss: Reflects the discriminator network's effectiveness in distinguishing between real and synthesized speech during training, indicating the network's ability to discern natural speech from artificially generated speech.

Generator Loss: Measures the success of the generator network in producing realistic speech outputs, demonstrating its capability to deceive the discriminator by generating natural-sounding speech.

Speech Synthesis Accuracy: Represents the accuracy of the text-to-speech system in converting input text into speech, providing insights into the system's effectiveness in preserving semantic meaning and linguistic nuances.

User Satisfaction Metrics: User feedback surveys and subjective evaluations assess user satisfaction with the synthesized speech outputs, considering factors such as naturalness, intelligibility, and overall listening experience.

Demo Link:
https://github.com/Dharshan-11/IBM-PROJECT-Gen-AI.git