A Synopsis

of

Minor Project [CC3270]

**Real-Time Detection of AI-Generated Speech**

Submitted to Manipal University Jaipur

towards the partial fulfillment of the requirement for the award of the degree of

**BACHELOR OF TECHNOLOGY**

in

Computer and Communication Engineering

Jan – May 2025

By:

Vansh Giri

229303045



Under the guidance of:

Dr. Vaishali Yadav and Dr. Arpit Kumar Sharma

**Department of Computer and Communication Engineering**

**Manipal University Jaipur**

**INTRODUCTION**

The rapid advancements in generative AI have led to significant breakthroughs in the speech domain, enabling real-time voice cloning and conversion. While these technologies have beneficial applications in entertainment, accessibility, and communication, they also pose serious ethical and security risks. AI-generated speech can be misused for impersonation, fraud, and misinformation, making it crucial to develop reliable detection mechanisms.

This project focuses on real-time detection of AI-generated speech used in deepfake voice conversions. With the rise of sophisticated deep learning models, distinguishing between real and synthetic voices has become increasingly challenging. The ability to detect such manipulations is essential for maintaining trust in digital interactions, safeguarding privacy, and preventing identity fraud.

To address this issue, I have chosen a dataset from Kaggle that includes both raw audio samples and pre-extracted features in CSV format. The dataset consists of real human speech and AI-converted voices, allowing for effective model training and evaluation. By leveraging advanced machine learning techniques, this project aims to develop a system capable of identifying AI-generated speech with high accuracy and efficiency.

A key aspect of this project is its real-time capability, which is critical for applications such as voice authentication, media verification, and cybersecurity. Additionally, the project will culminate in the development and hosting of a web-based tool that can analyse voice inputs and determine their authenticity. This user-friendly interface will make AI voice detection accessible to a broader audience, including individuals, organizations, and law enforcement agencies.

By choosing this as my minor project, I aim to contribute to the growing field of AI security and ethical AI usage. The increasing prevalence of deepfake technology makes it imperative to develop countermeasures that can help detect and mitigate potential threats. This project will not only enhance my understanding of machine learning and deepfake detection but also address a real-world challenge with significant societal implications.

**MOTIVATION**

The widespread availability of AI-generated voice technology has introduced both opportunities and threats to digital communication. While synthetic voices can enhance accessibility and content creation, they also pose risks related to deception, fraud, and privacy breaches. The misuse of AI-generated speech in deepfake scams and misinformation campaigns underscores the urgent need for robust detection mechanisms.

My motivation for this project stems from a strong interest in AI ethics and cybersecurity. As deepfake voice technology becomes more advanced, it is essential to develop real-time solutions that can detect and counteract fraudulent uses. This project provides an opportunity to work with state-of-the-art machine learning models, gain hands-on experience in audio signal processing, and contribute to a pressing global issue.

Furthermore, the challenge of building a real-time AI voice detection system aligns with my goal of applying theoretical knowledge to practical, impactful solutions. The development of a web-based detection tool ensures that this research can have real-world applications, making it accessible to individuals and organizations that need to verify the authenticity of voice recordings.

By working on this project, I aim to bridge the gap between AI advancements and security concerns, ensuring that technological progress does not come at the cost of ethical risks. This project is not only a technical endeavour but also a step towards making AI safer and more responsible.

**PROBLEM STATEMENT**

With the rise of AI-driven speech synthesis and deepfake voice generation, there is an increasing risk of malicious actors using these technologies for fraudulent activities such as identity theft, misinformation, and scams. Current detection mechanisms are either ineffective in real-time scenarios or require significant computational resources, making them impractical for widespread deployment.

The primary challenge is to develop a real-time AI voice detection system that can accurately distinguish between human speech and AI-generated speech. The solution should be scalable, efficient, and accessible, ensuring that individuals and organizations can verify the authenticity of voice recordings without requiring specialized expertise.

**Objectives :**

1. Develop a machine learning-based model capable of detecting AI-generated speech in real-time.
2. Utilize a dataset comprising real and fake audio samples to train and evaluate the model.
3. Implement a web-based tool that allows users to analyse voice recordings for deepfake detection.
4. Ensure the model is lightweight and optimized for real-time performance.
5. Assess the model’s accuracy, efficiency, and robustness against evolving deepfake techniques.

**Pros & Cons of Existing Methods:**

**Pros :**

* Traditional spectral analysis and acoustic feature extraction methods offer some effectiveness in differentiating real and AI-generated speech.
* Deep learning approaches, such as convolutional and recurrent neural networks, have shown promising results in detecting synthetic audio.
* Existing forensic tools can analyse speech patterns, pitch variations, and inconsistencies to flag potential deepfake voices.

**Cons :**

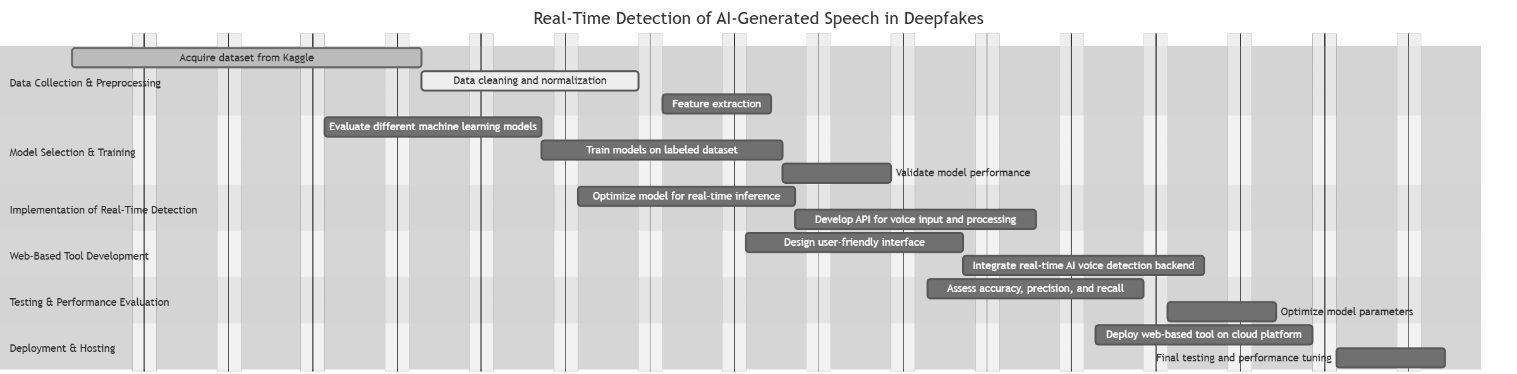
* Many current detection systems are computationally expensive and require powerful hardware.
* Some models rely heavily on supervised learning, making them vulnerable to adversarial attacks and generalization issues.
* The evolving nature of AI-generated speech means that detection methods must continuously adapt to new synthesis techniques.
* Lack of publicly available real-time detection tools makes it challenging for non-experts to verify voice authenticity.

By addressing these limitations, this project aims to create a practical and effective solution for real-time AI voice detection, contributing to the broader effort of safeguarding digital communications from deepfake threats.

**METHODOLOGY**

**Methodology:** To achieve the objectives of this project, the following steps will be followed:

1. **Data Collection & Preprocessing**
   * Acquire the dataset from Kaggle, containing real and AI-generated speech.
   * Perform data cleaning, feature extraction, and normalization.
2. **Model Selection & Training**
   * Evaluate various machine learning models (CNN, LSTM, Transformer-based architectures).
   * Train models using labelled datasets and validate performance with test data.
3. **Implementation of Real-Time Detection**
   * Optimize model for real-time inference with minimal latency.
   * Develop an API for voice input and processing.
4. **Web-Based Tool Development**
   * Design and implement a user-friendly web interface for voice analysis.
   * Integrate real-time AI voice detection with backend processing.
5. **Testing & Performance Evaluation**
   * Assess the accuracy, precision, and recall of the detection system.
   * Optimize model parameters for improved real-time performance.
6. **Deployment & Hosting**
   * Deploy the web-based detection tool on a cloud platform.
   * Conduct final testing and performance tuning before release.



Gantt Chart of Methodology

This structured methodology ensures a systematic approach towards achieving the project's objectives and delivering a functional AI voice detection system.

**FACILITIES REQUIRED**

**Software Requirements:**

* Python (with libraries such as TensorFlow, PyTorch, Librosa, and Scikit-learn)
* Flask/Django for web backend development
* React.js/HTML/CSS for frontend development
* MongoDB/MySQL for database management (if required)
* Cloud services (AWS, Google Cloud, or Azure) for hosting
* Jupyter Notebook for development and experimentation

**Hardware Requirements:**

* There are no specific hardware requirements for this project, as it will be conducted on my personal computer. In the event that additional computational resources are needed, I will request access to the lab's equipment.

**BIBLIOGRAPHY & REFERENCES**

Dataset: [DEEP-VOICE: DeepFake Voice Recognition](https://www.kaggle.com/datasets/birdy654/deep-voice-deepfake-voice-recognition)

Research Papers:

* [A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions](https://www.mdpi.com/1999-4893/15/5/155)
* [Detecting Deepfake Voice Using Explainable Deep Learning Techniques](https://www.mdpi.com/2076-3417/12/8/3926)