DonFakeLah!:

Detecting Deepfake Audio

DSI-SG-42
Gilbert

TABLE OF CONTENTS

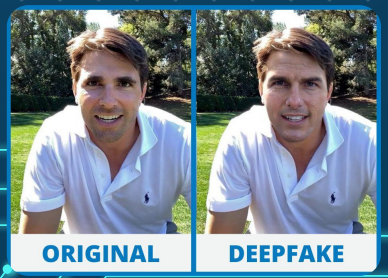
01

Introduction

Growth of AI on Media

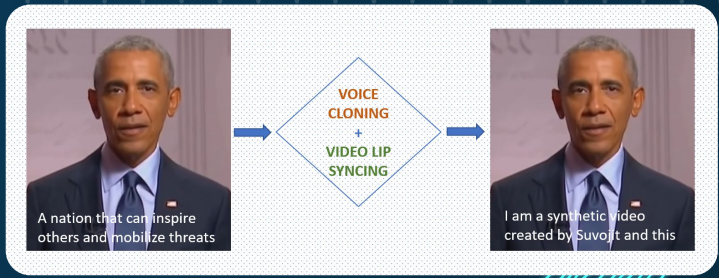
Synthetic Media

Image



Audio

Video



Pros - Its Application

Voice Assistant

GOOGLE

SIRI

ALEXA

Content Creators

INSTAGRAM

TIKTOK

YOUTUBE

Customer Service

Cons - How It’s Misused

Source: BBC News Jan 2024

Source: CNN News Feb 24

PM Lee warns against responding to deepfake videos of him promoting investment scams

Source: Straits Times Dec 2023

Deepfake video of DPM Lawrence Wong promoting investment scam circulating on social media

Source: Straits Times Dec 2023

WHAT IF?

They Impersonate

Family

Friends

Loved Ones

Who is Jack?

A 30 years old ambitious man, determined to achieve financial independence and retire early.

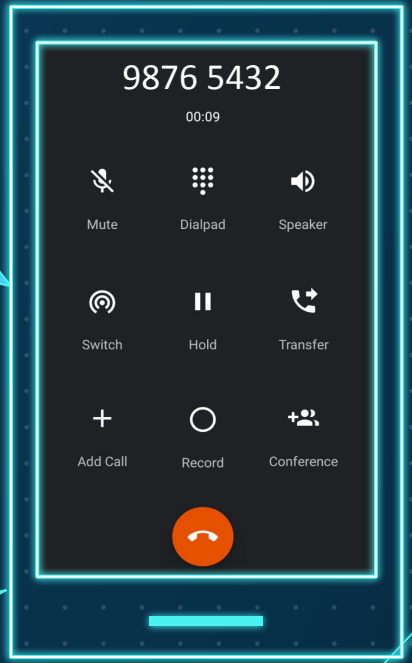
Jack has been actively building his investment portfolio and diligently saving for the future. He frequently exchange investment information with his good friend Jackson.

What Happened to Jack?

Hey Jack, I've got this amazing investment opportunity that I wanted to share with you. It's a sure thing, guaranteed to triple your money in no time. All you need to do is transfer some funds to me, and I'll take care of the rest.

Wah! let’s go for it bro

I'll give you the details right away. Thanks bro. You won't regret it.



Hey Jackson, how's the investment going? Any news?

??? What are you talking about? I didn't call you about any investment. Are you sure it was me?

Oh no... I think I've been scammed. My savings 😭

Is there any way to determine whether the voice I am hearing is real or not?

I hope that I can protect myself and my family from this kind of scam

02

Problem Statement

“

How can we develop a model to effectively detect **deepfake audio recordings**, distinguishing between **genuine human speech and AI generated sound** for ensuring audio authenticity and combating the spread of misinformation and fraudulent activities?

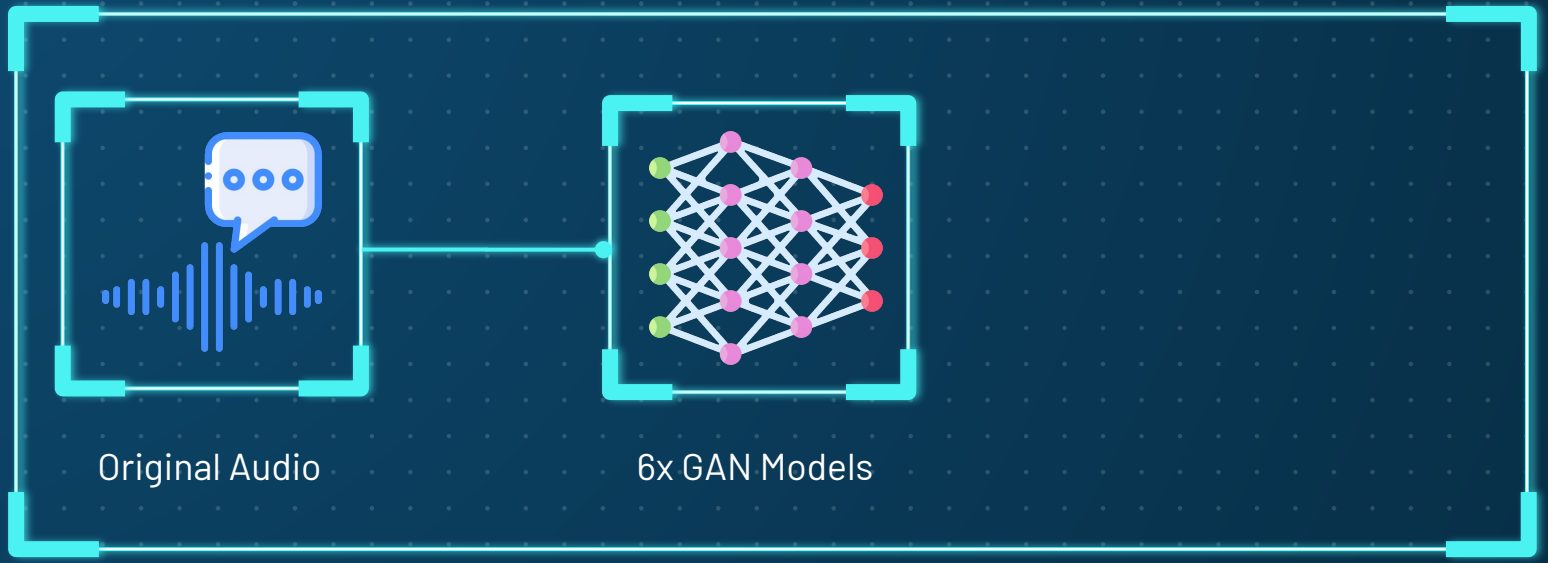
03

Audio Analysis

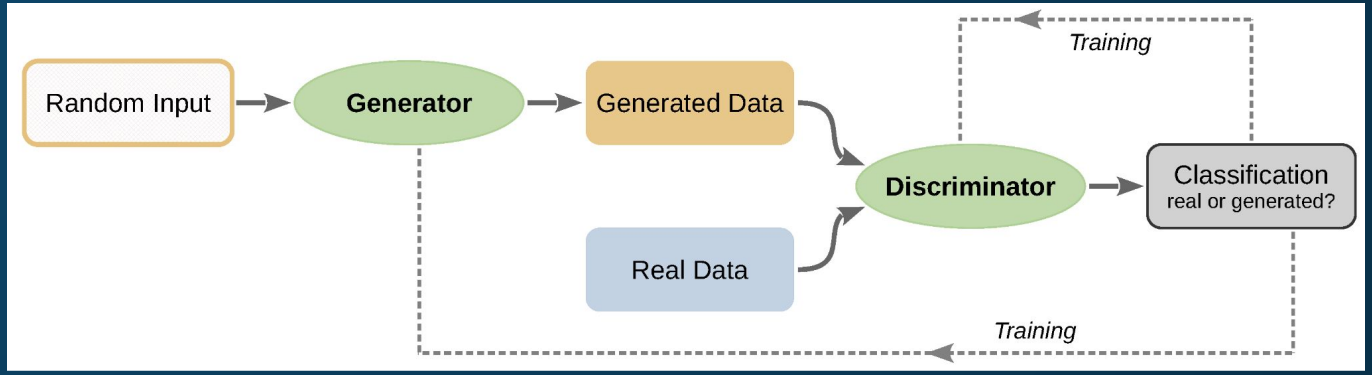
Dataset Source

- LJ Speech 1.1
  - Short audio clips of a single speaker reading passages from 7 non-fiction books
  - Total: 13,100 audio clips
  - Label: Original Audio

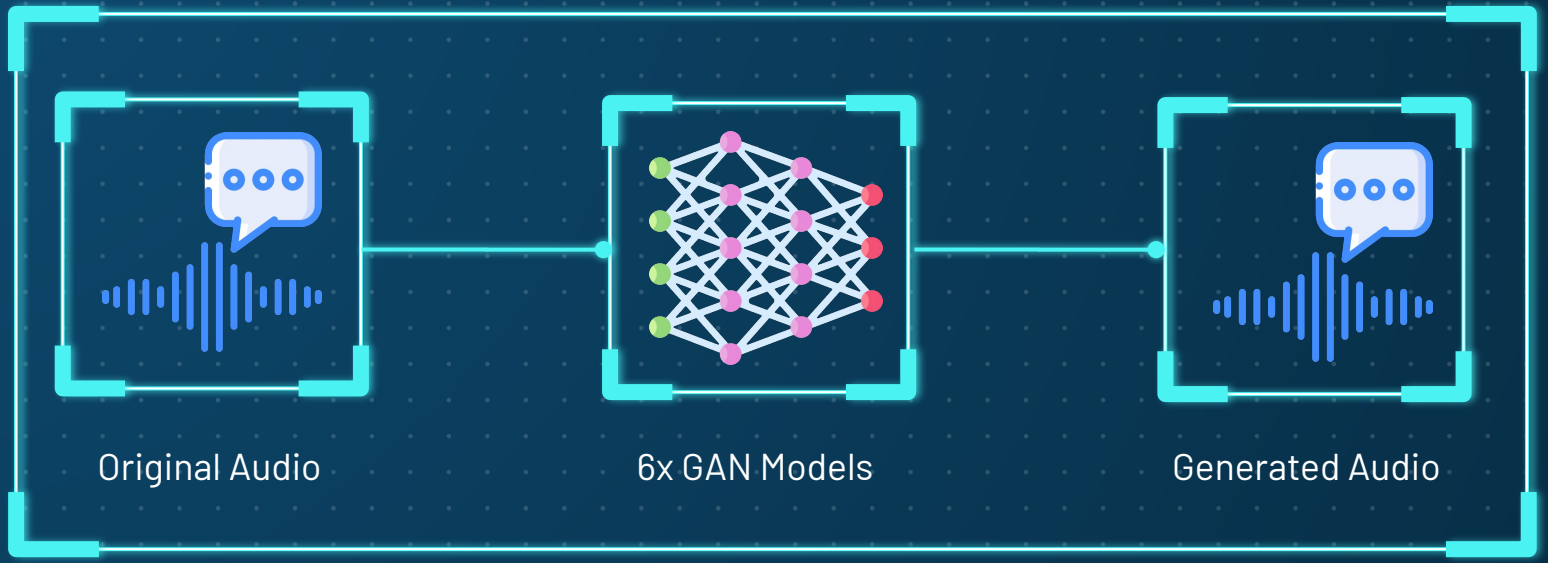
Dataset Overview



What is Generative Adversarial Network (GAN)?



Dataset Overview



Original Audio	13,100
Generated Audio	78,600

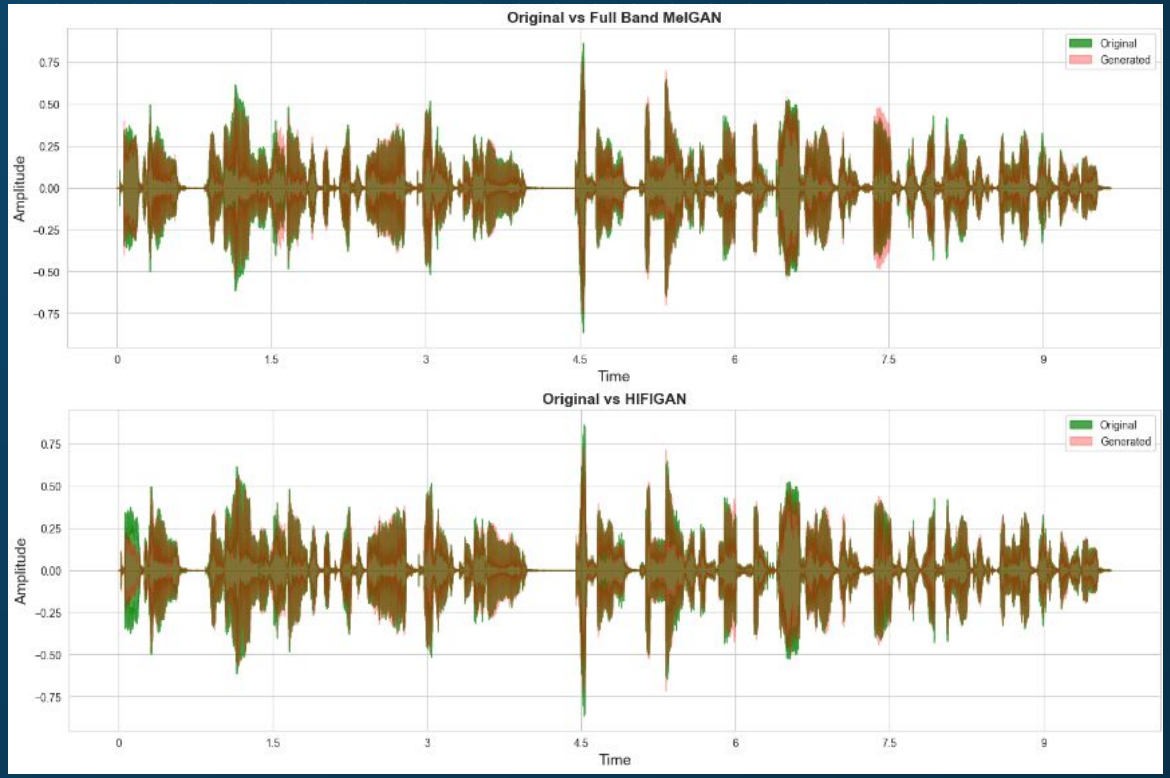
Real vs Fake – How Does It **Sound** Like

REAL

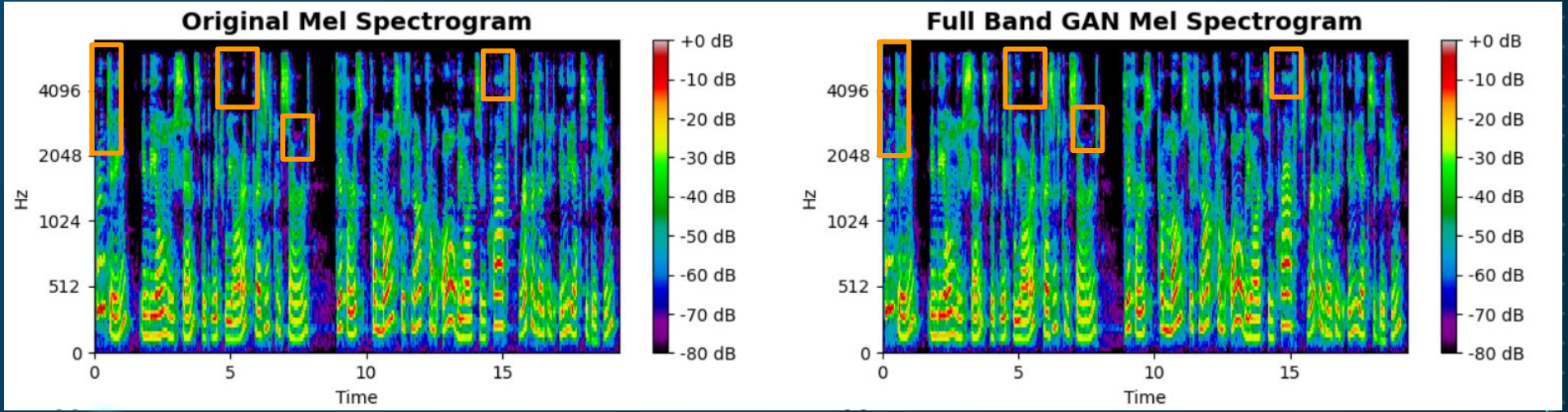
FAKE

FAKE

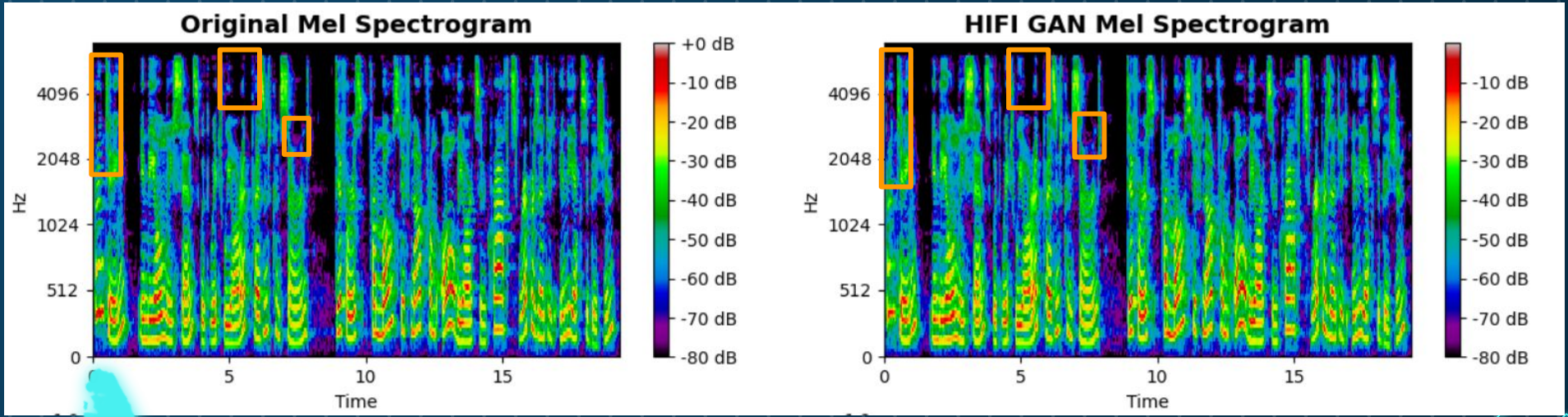
Real vs Fake – How Does It **Look** Like?



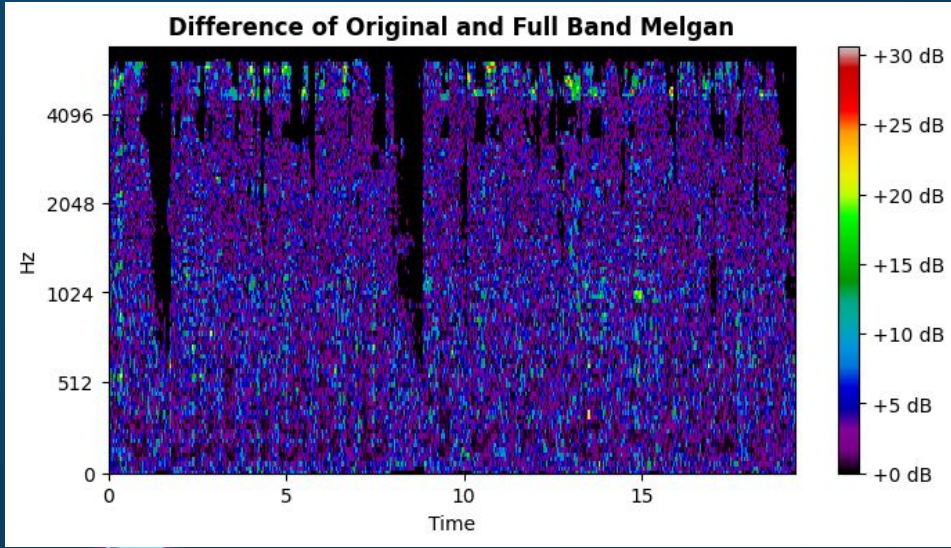
Original vs Full Band GAN Mel Spectrogram

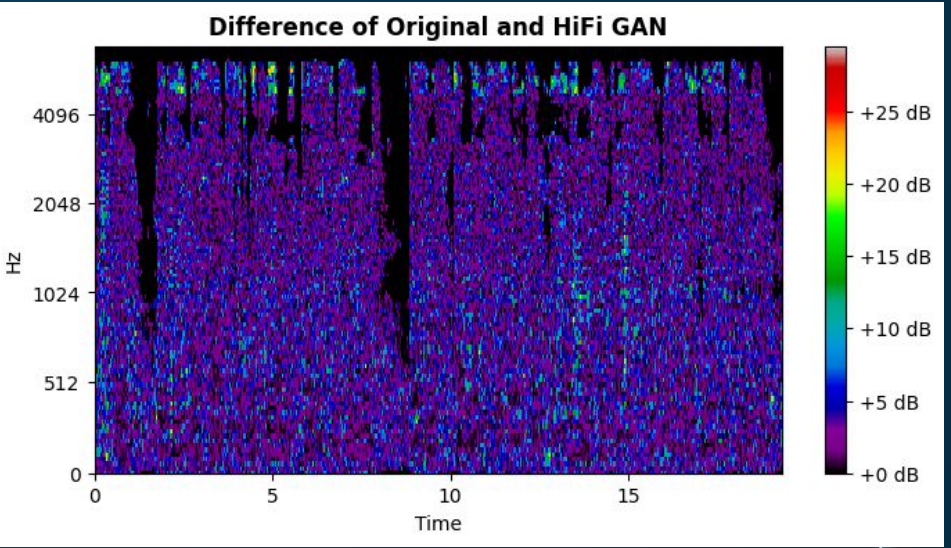


Original vs HiFI GAN Mel Spectrogram



Original vs GANs - Difference

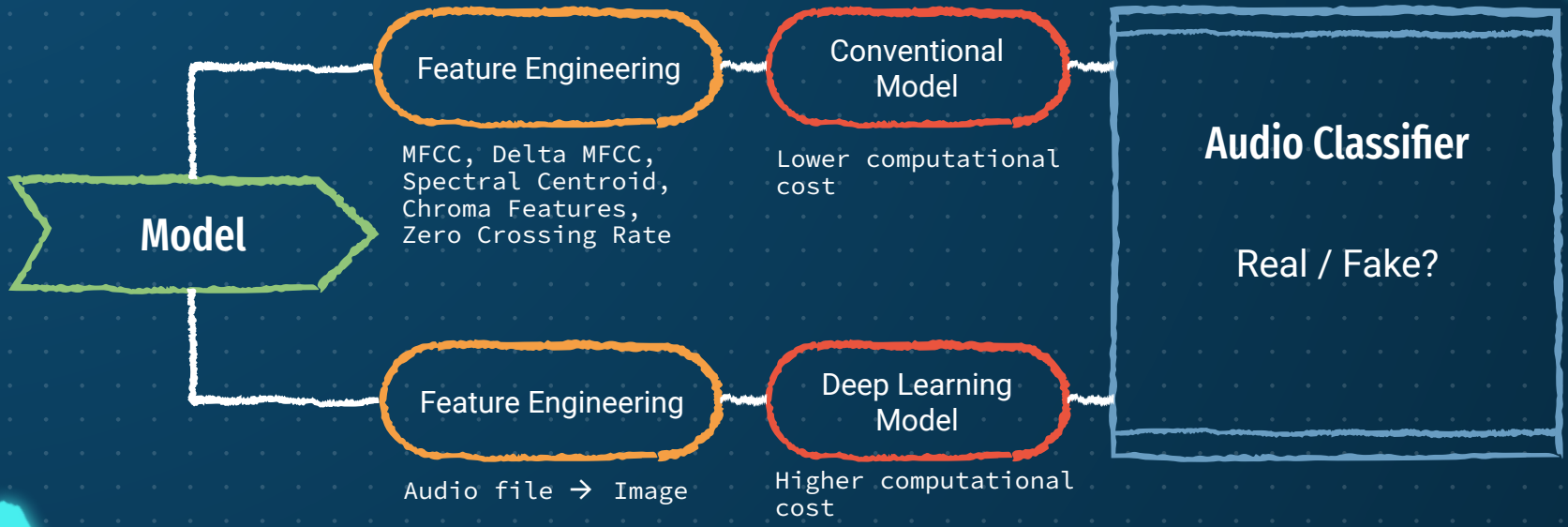




04

Model Development

Machine Learning Model



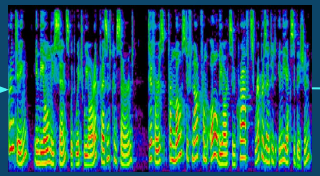
Conventional Method: Evaluating Baseline Model

Metrics / Model	Accuracy		Recall	
	Train	Test	Train	Test
Random Forest	1.000	0.726	1.000	0.825
Logistic Regression	0.688	0.674	0.674	0.673
Decision Tree	1.000	0.622	1.000	0.672
ADABoost	0.665	0.632	0.644	0.632
XGBoost	0.834	0.696	0.810	0.735
SVC	0.835	0.726	0.804	0.766

Conventional Method: Hypertuned model

Metrics / Model	Model Stages	Accuracy		Recall	
		Train	Test	Train	Test
Random Forest	Baseline	1.000	0.726	1.000	0.825
	Hypertuned	0.749	0.632	0.780	0.655
Logistic Regression	Baseline	0.688	0.674	0.674	0.673
	Hypertuned	0.688	0.676	0.689	0.676
XGBoost	Baseline	0.834	0.696	0.810	0.735
	Hypertuned	0.857	0.690	0.869	0.732

Deep Learning – CNN + LSTM

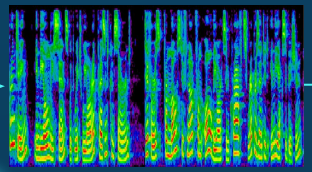


Mel Spectrogram Image

CNN

LSTM

Real/Fake



Mel Spectrogram Image

CNN

LSTM

Original or one of the generators

Summary of Model Evaluation

Metrics / Model	Accuracy		Recall	
	Train	Test	Train	Test
Random Forest	0.749	0.632	0.780	0.655
Logistic Regression	0.688	0.676	0.689	0.676
XGBoost	0.857	0.690	0.869	0.732
CNN + LSTM (Binary Classification)	0.908	0.857	0.980	1.000
CNN + LSTM (Multi-Class Classification)	0.970	0.929	0.967	0.928

05

Implementation

Problem Statement:

How can we develop a model to effectively detect **deepfake audio recordings**, distinguishing between **genuine human speech and AI generated sound** for ensuring audio authenticity and combating the spread of misinformation and fraudulent activities?

Is there an app to help Jack to identify if the audio is real or fake?

App Demo

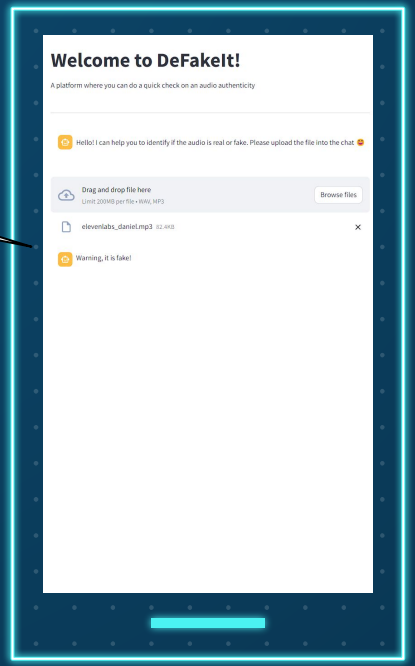
Audio Sample

Recorded Voice

Elevenlabs

With **DonFakeLah**,
now Jack can identify if the audio is fake / real

Warning, it is fake!

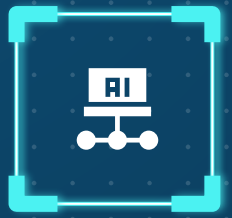


HA! You are fake! Nice try 😡

06

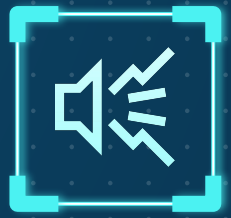
Conclusion

Limitations



Unique GAN Architecture

- Model might not perform well to detect new type of audio generator



Audio Noises

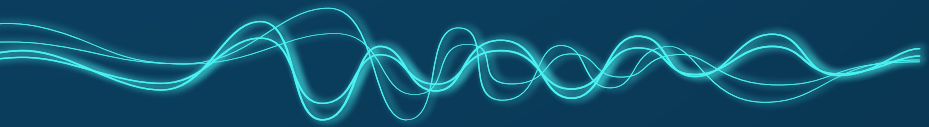
- Background noise from the environment

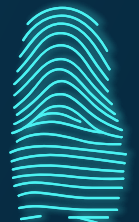
Audio Volume

- Different volume affect model detectability

Conclusion

- The model developed could distinguish original or generated audio
- The model aims to help the user identify generated audio to protect them from scam and misinformation





Future Works

- Future application
  - Implement to SPF website
  - Collaborate with Smartphone providers for model implementation
- Addressing limitations:
  - Collect data from new architecture and train to the model
  - Train model with audio files with noises and different volumes

THANKS !