This caselet was developed by student Kavin Manivannan as a resource for Community Infrastructure to Strengthen AI for Audio Deepfake Analysis (CISAAD), a project funded by the National Science Foundation Award #2346473 and **supervised by** Dr. Vandana P. Janeja, Dr. Christine Mallinson **and Dr. Karen Chen** at the University of Maryland, Baltimore County (UMBC). For more information, visit cisaad.umbc.edu or email cisaad@umbc.edu.

Muting Misinformation as a Deepfake Detective

In this caselet, we'll explore audio deepfakes through a real world challenge! Read the background information carefully, and use it to answer the multiple choice questions that follow.

By the end of this caselet, you will become a **Deepfake Detective.** You'll know how to spot audio deepfakes on your own, helping you stay more vigilant and protected in your everyday online activities. Good luck!

* Indicates required question

Picture This!

Think of your favorite celebrity. Now imagine you're scrolling on social media, and you come across a video of them. You hear them say something outrageous! It’s going viral—thousands of people are sharing it, commenting, and freaking out.

Here’s the twist… it’s **completely fake.**

You and thousands of people have just been the victim of an **audio deepfake.**

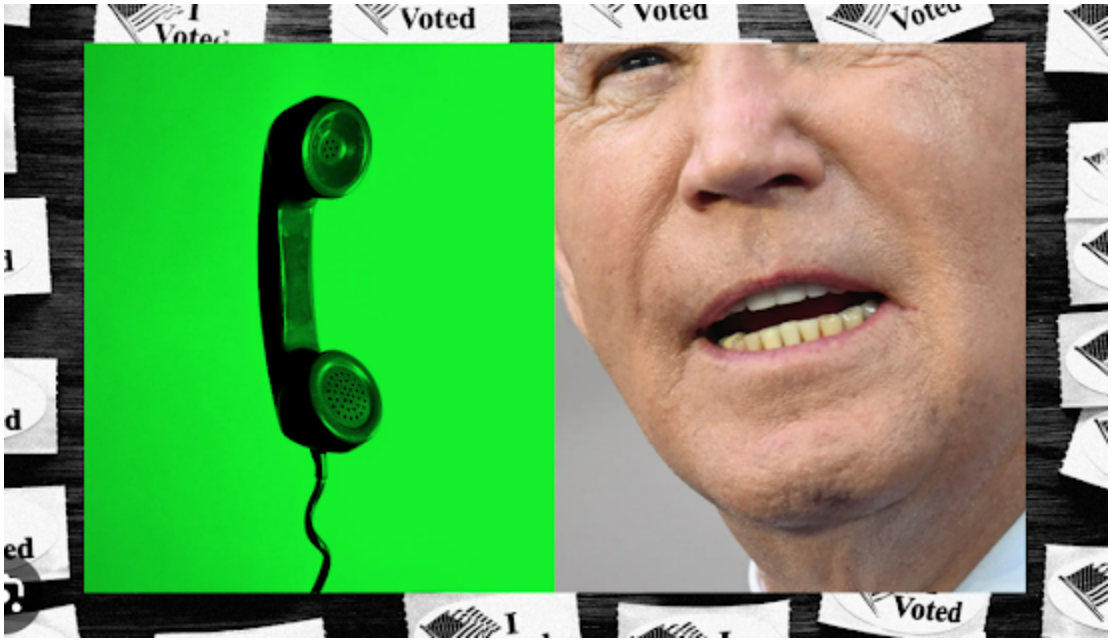
What is an audio deepfake?

An audio deepfake is a fake voice recording created by AI to sound exactly like someone else. Think of it like a high-tech impersonation, where a computer makes it seem like a person is saying something they never actually said. People might use this to trick others, spread lies, or pretend to be someone they're not.

This caselet was developed by student Kavin Manivannan as a resource for Community Infrastructure to Strengthen AI for Audio Deepfake Analysis (CISAAD), a project funded by the National Science Foundation Award #2346473 and **supervised by** Dr. Vandana P. Janeja, Dr. Christine Mallinson **and Dr. Karen Chen** at the University of Maryland, Baltimore County (UMBC). For more information, visit cisaad.umbc.edu or email cisaad@umbc.edu.

Curious to hear an audio deepfake? Click Here!

Problem Context



This caselet was developed by student **Kavin Manivannan as a resource for Community Infrastructure to Strengthen AI for Audio Deepfake Analysis (CISAAD), a project funded by the National Science Foundation Award #2346473 and supervised by** Dr. Vandana P. Janeja, Dr. Christine Mallinson **and Dr. Karen Chen** at the University of Maryland, Baltimore County (UMBC). For more information, visit cisaad.umbc.edu or email cisaad@umbc.edu.

The clip you just heard was of the 46th President Joe Biden. **Or was it?** This audio deepfake of Joe Biden was sent to voters in New Hampshire. The deepfake audio told voters to not vote in an upcoming primary election. “Your vote makes a difference in November, not this Tuesday,” the voice mimicking Biden says.

Deepfake audio clips like these can invade social media and sway public opinion with misinformation and undermine public trust in the news.

To solve this problem, it's time to think like a data scientist. Imagine you're a data scientist that has been hired by a prestigious news network to build an AI (*artificial intelligence*) system to detect deepfake audio clips. This news channel prides itself on delivering accurate and reliable news to millions of people.

Caselet Questions

1. **Question 1.** 1 point

   The news agency wants to understand *how* your system can tell if a voice is real or fake. What kind of system will you create to make it easier for them to understand?

   Mark only one oval.

   - ◯ Build a simple system that's easy to explain
   - ◯ Build a more complex system that's harder to explain but might catch more deepfakes

Question 2:

Let’s explore some tools that could help us spot fake audio clips. Different tools work in different ways. We want to find one that’s not only accurate but also easy to explain.

This caselet was developed by student Kavin Manivannan as a resource for Community Infrastructure to Strengthen AI for Audio Deepfake Analysis (CISAAD), a project funded by the National Science Foundation Award #2346473 and **supervised by** Dr. Vandana P. Janeja, Dr. Christine Mallinson **and Dr. Karen Chen** at the University of Maryland, Baltimore County (UMBC). For more information, visit cisaad.umbc.edu or email cisaad@umbc.edu.

Model	Definition
Logistic Regression	A simple system that helps sort things into two categories, like real or fake.
CNN	A system that looks at patterns in things like audio waves, trying to spot fake parts.
Decision Tree	Breaks the data down into easy steps, making decisions along the way to see if it's real or fake.
GAN	*A challenging system where two programs compete—one makes fake things, and the other tries to figure out if it's fake.*
Random Forest	Uses lots of decision trees to figure out if something is fake, especially when the data is noisy.

2. Given the goal of explaining how we detect deepfakes, which tools would you consider using? *(Check all that apply)* *

Check all that apply.

- [ ] Logistic Regression
- [ ] CNN
- [ ] Decision Tree
- [ ] GAN
- [ ] Random Forest

Question 3

There are many different deepfake types you may come across.

This caselet was developed by student Kavin Manivannan as a resource for Community Infrastructure to Strengthen AI for Audio Deepfake Analysis (CISAAD), a project funded by the National Science Foundation Award #2346473 and **supervised by** Dr. Vandana P. Janeja, Dr. Christine Mallinson **and Dr. Karen Chen** at the University of Maryland, Baltimore County (UMBC). For more information, visit cisaad.umbc.edu or email cisaad@umbc.edu.

Attack Type	Definition
Text-to-Speech	Synthesizing speech from written text using machine learning models.
Voice Conversion	Synthetically transforms one person's voice into another's using machine learning models.
Mimicry	Impersonates another person's voice by analyzing clips of their actual voice.
Replay Attacks	Playback of a real person's voice.

Listen to each type below:

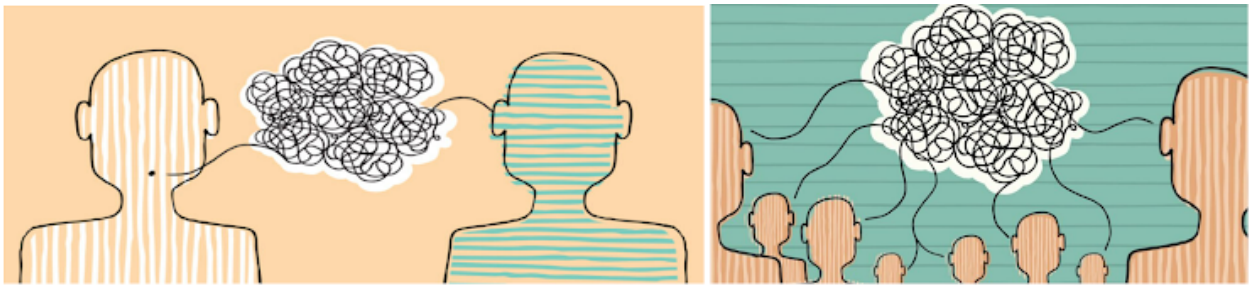
Text-to-Speech
Voice Conversion
Mimicry
Replay Attacks

3. Based on the different types of deepfakes in the chart, which types do you think are the hardest to detect and why? 1 point

This caselet was developed by student Kavin Manivannan as a resource for Community Infrastructure to Strengthen AI for Audio Deepfake Analysis (CISAAD), a project funded by the National Science Foundation Award #2346473 and **supervised by** Dr. Vandana P. Janeja, Dr. Christine Mallinson **and Dr. Karen Chen** at the University of Maryland, Baltimore County (UMBC). For more information, visit cisaad.umbc.edu or email cisaad@umbc.edu.

Question 4:

Now, how are we actually going to detect deepfakes? We know what we need to fix, and we have our toolbox. So, what tools are we going to use?



This caselet was developed by student Kavin Manivannan as a resource for Community Infrastructure to Strengthen AI for Audio Deepfake Analysis (CISAAD), a project funded by the National Science Foundation Award #2346473 and **supervised by** Dr. Vandana P. Janeja, Dr. Christine Mallinson **and Dr. Karen Chen** at the University of Maryland, Baltimore County (UMBC). For more information, visit cisaad.umbc.edu or email cisaad@umbc.edu.

Read Carefully

As you dive into your research, you stumble upon insights from the world of *variationist sociolinguistics*—a branch of linguistics that explores how language shifts across different social contexts. Imagine how we speak, not as a rigid structure, but as a dynamic, ever-changing expression, influenced by who we're talking to, where we are, and even how we feel. Human language is full of these subtle variations; we tweak our tone, pitch, and pronunciation in ways we often don't even notice. These phonetic nuances are a complex dance of sounds, both deliberate and intuitive, that reflect the diversity of our communication.

You contact a sociolinguistic expert at a local university, hoping they can offer insights on your dataset to help you develop a new method for detecting deepfake speech. They're excited to help!

Here's what they found:

All genuine clips have *linguistic features* that help differentiate them from deepfake clips.

There are five of them:

Pitch – Higher or lower tone of a speech sample.

Pause – Break in the middle of speech production.

Initial/Final Bursts – Burst of air produced when beginning to pronounce a consonant, or at the end of pronouncing a consonant. If genuine, the bursts create the sounds */p/, /b/, /t/, /d/, /k/,* and */g/*.

Intake/Outtake of Breath – Breathing in or out.

Audio Quality – Overall audio quality of a sample. This includes any disturbances or distortion in the audio.

Let's put these features to use! Listen to this clip you heard previously → here. This time, rate whether you hear an anomalous feature for any of the five features, such as an unusually high pitch where there shouldn't be.

This caselet was developed by student Kavin Manivannan as a resource for Community Infrastructure to Strengthen AI for Audio Deepfake Analysis (CISAAD), a project funded by the National Science Foundation Award #2346473 and **supervised by** Dr. Vandana P. Janeja, Dr. Christine Mallinson **and Dr. Karen Chen** at the University of Maryland, Baltimore County (UMBC). For more information, visit cisaad.umbc.edu or email cisaad@umbc.edu.

4. **Part A** 1 point

How would you score this clip on its linguistic features? Give it a **1** if you hear an anomaly, or **0** if you don't.

Mark only one oval.

- ◯ Pitch: 0 Pause: 0 Burst: 1 Breath: 0 Audio Quality: 0
- ◯ Pitch: 0 Pause: 1 Burst: 0 Breath: 0 Audio Quality: 1
- ◯ Pitch: 1 Pause: 1 Burst: 0 Breath: 0 Audio Quality: 1
- ◯ Pitch: 1 Pause: 0 Burst: 0 Breath: 1 Audio Quality: 0

5. **Part B** 1 point

Based on your rating, is this clip genuine or an audio deepfake?

Mark only one oval.

- ◯ Genuine
- ◯ Deepfake

6. **Question 5:** * 0 points

Let's try one more! Listen to this clip → here.

Part A

How would you score this clip on its linguistic features? Give it a **1** if you hear an anomaly, or **0** if you don't.

Mark only one oval.

- ◯ Pitch: 1 Pause: 0 Burst: 1 Breath: 0 Audio Quality: 0
- ◯ Pitch: 0 Pause: 1 Burst: 1 Breath: 0 Audio Quality: 1
- ◯ Pitch: 0 Pause: 1 Burst: 1 Breath: 0 Audio Quality: 0
- ◯ Pitch: 0 Pause: 1 Burst: 0 Breath: 0 Audio Quality: 0

This caselet was developed by student Kavin Manivannan as a resource for Community Infrastructure to Strengthen AI for Audio Deepfake Analysis (CISAAD), a project funded by the National Science Foundation Award #2346473 and **supervised by** Dr. Vandana P. Janeja, Dr. Christine Mallinson **and Dr. Karen Chen** at the University of Maryland, Baltimore County (UMBC). For more information, visit cisaad.umbc.edu or email cisaad@umbc.edu.

7. **Part B** 1 point

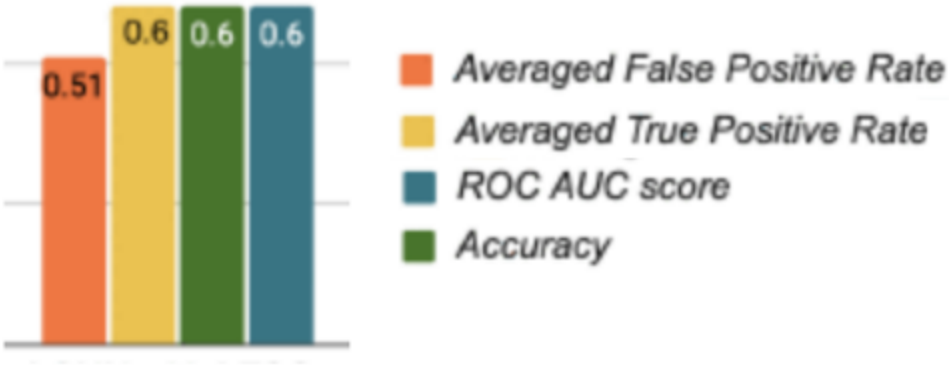
Based on your rating, is this clip genuine or an audio deepfake?

Mark only one oval.

- ◯ Genuine
- ◯ Deepfake

Question 6:

Now we're ready to test! We measure performance by four metrics. The graph below details the metrics below:



8. Assume identifying an audio clip as a deepfake is positive. 1 point

What is a false positive in this context?

Mark only one oval.

- ◯ When the system correctly identifies a deepfake as a fake.
- ◯ When the system fails to detect a deepfake, it classifies it as genuine.
- ◯ When the system finds a genuine audio clip, and classifies it as genuine.
- ◯ When the system incorrectly identifies a genuine audio clip as a deepfake.

This caselet was developed by student Kavin Manivannan as a resource for Community Infrastructure to Strengthen AI for Audio Deepfake Analysis (CISAAD), a project funded by the National Science Foundation Award #2346473 and **supervised by** Dr. Vandana P. Janeja, Dr. Christine Mallinson **and Dr. Karen Chen** at the University of Maryland, Baltimore County (UMBC). For more information, visit cisaad.umbc.edu or email cisaad@umbc.edu.

9. **Question 7:** 1 point

Assume identifying an audio clip as a deepfake is positive.

What is the cost of error if a false positive occurs?

Mark only one oval.

- ◯ Loss of revenue due to blocked legitimate content.
- ◯ Damage to reputation due to falsely accusing individuals or organizations.
- ◯ Increased vulnerability to deepfake attacks due to system compromise.
- ◯ None of the above

Credits

This caselet was created by Kavin Manivannan

The data used in this caselet is adapted from Baseline_Master_Data_SampleSet3.xlsx, under the IEEE License

Citations:

Z. Khanjani, L. Davis, A. Tuz, K. Nwosu, C. Mallinson and V. P. Janeja, "Learning to Listen and Listening to Learn: Spoofed Audio Detection Through Linguistic Data Augmentation," 2023 IEEE International Conference on Intelligence and Security Informatics (ISI), Charlotte, NC, USA, 2023, pp. 01-06, doi: 10.1109/ISI58743.2023.10297267. keywords: {Training;Annotations;Linguistics;Phonetics;Media;Feature extraction;Security;audio deepfake;spoofed audio detection;Artificial Intelligence;linguistics;sociolinguistics;linguistic perception},

Swenson, Ali, and Will Weissert. “New Hampshire investigating fake Biden robocall meant to discourage voters ahead of primary.” *Associated Press*, 24 Jan. 2024, https://apnews.com/article/new-hampshire-primary-biden-ai-deepfake-robocall-f3469ceb6dd613079092287994663db5.

Image Credits:

Image 1 (Header): Getty / NurPhoto
Image 2: WaveVisual

Image 3: WIRED