IAFIS Biometrics & Forensics

Suplantación de Identidad a Través de la IA

Riesgos, Casos Reales y Estrategias para la Protección de Entidades Gubernamentales

El nuevo panorama de la clonación de voz

- 2016, Star Wars Rogue One
  Uso de la imagen y voz del fallecido actor **Peter Cushing**

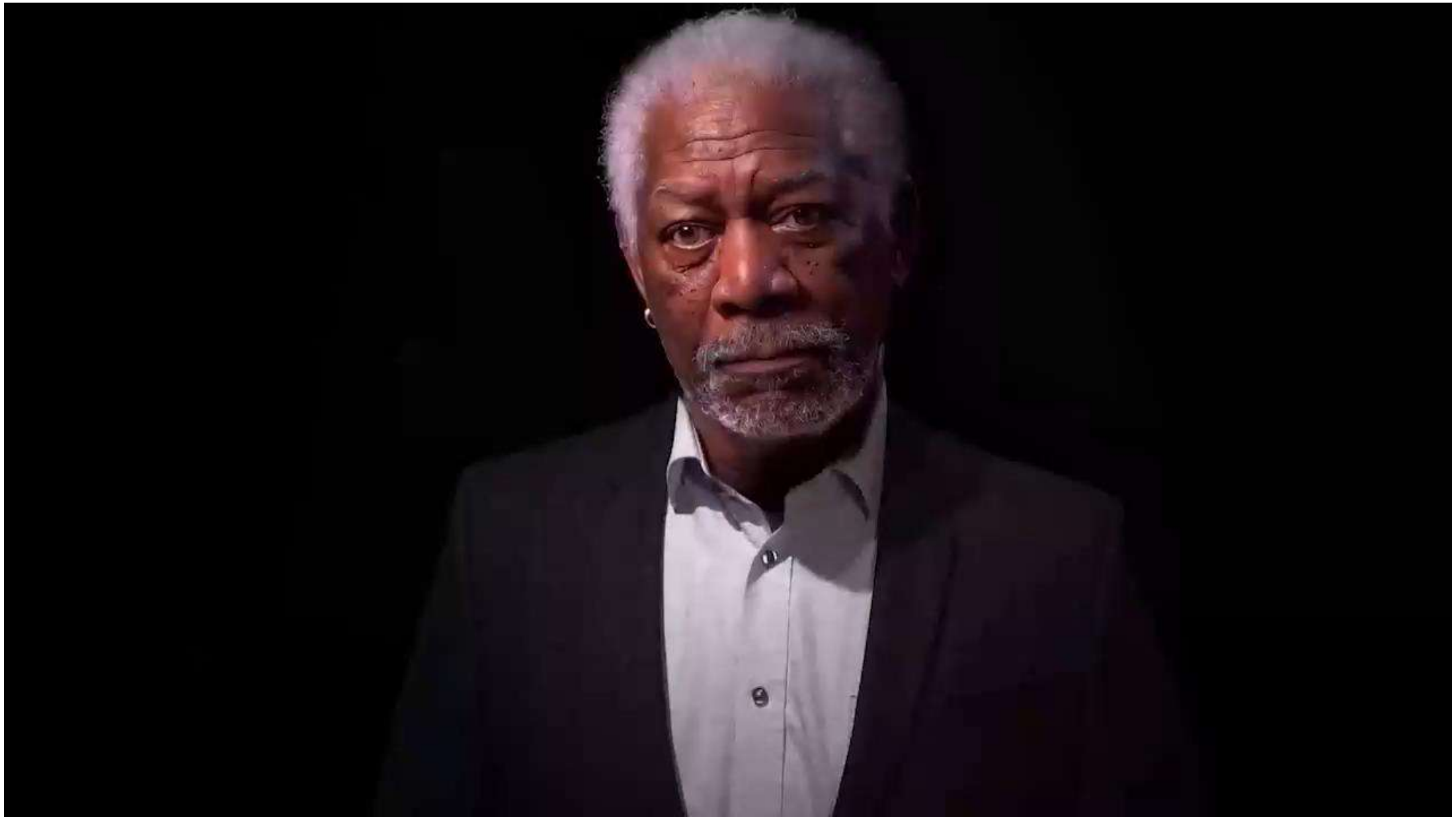
- Avance en la producción cinematográfica y muestra del potencial de la tecnología.

Preguntas Críticas

¿Qué pasa si un criminal utuliza esta tecnología?
¿Estamos preparados para enfrentar el uso malicioso?

Algunos ejemplos de esta tecnología


[Música]
Déjame hablarte




AI

PEOPLE ARE POORLY EQUIPPED TO DETECT AI-POWERED VOICE CLONES

A PREPRINT

Sarah Barrington
School of Information
University of California, Berkeley
Berkeley, CA, 94720
sbarrington@berkeley.edu

Hany Farid
School of Information
Department of Electrical Engineering
and Computer Sciences
University of California, Berkeley
Berkeley, CA, 94720
hfarid@berkeley.edu

October 8, 2024

ABSTRACT

As generative AI continues its ballistic trajectory, everything from text to audio, image, and video generation continues to improve in mimicking human-generated content. Through a series of perceptual studies, we report on the realism of AI-generated voices in terms of identity matching and naturalness. We find human participants cannot reliably identify short recordings (less than 20 seconds) of AI-generated voices. Specifically, participants mistook the identity of an AI-voice for its real counterpart 80% of the time, and correctly identified a voice as AI-generated only 60% of the time. In all cases, performance is independent of the demographics of the speaker or listener.

Keywords: Generative AI | Voice Cloning | Voice Identification

Más ejemplos...

Caso reciente con artistas contemporaneos

Preguntas Críticas

Si es possible recrear a un artista,
¿qué impide clonar la voz de un funcionario público?


BBC
NEWS
.GOV

Caso en Reino Unido en 2019

- Los delincuentes utilizaron software basado en inteligencia artificial para hacerse pasar por la voz de un director ejecutivo y exigir una transferencia fraudulenta de 220.000 euros (243.000 dólares)

WSJ PRO CYBERSECURITY

Hogar Noticia Archivo de investigación Boletines Eventos

WSJ PRO

Los estafadores utilizaron la IA para imitar la voz del CEO en un inusual caso de ciberdelincuencia

Las estafas con inteligencia artificial son un nuevo reto para las empresas

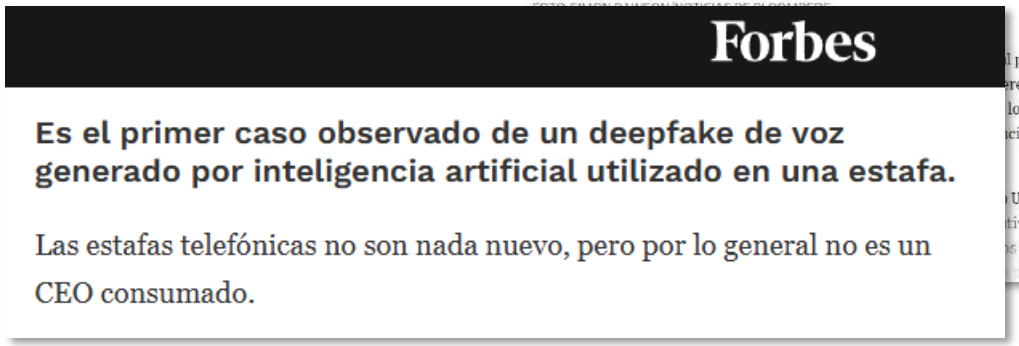
Por Catherine Stupp

Actualizado el 30 de agosto de 2019 a las 12:52 pm ET | WSJ PRO

Compartir Redimensionar

LECTURAS OBLIGADAS DE CIBERSEGURIDAD

1. El puerto de Róterdam pone a prueba la red cuántica para defenderse de los hackeos
2. La actualización del hackeo de Okta muestra los desafíos de las divulgaciones cibernéticas rápidas
3. ¿Cómo están construyendo la cultura corporativa en tiempos híbridos?
4. El sector eléctrico europeo carece de expertos cibernéticos a medida que la guerra de Ucrania aumenta los riesgos de piratería
5. La conciliación de las normas internacionales de notificación de



Forbes

Es el primer caso observado de un deepfake de voz generado por inteligencia artificial utilizado en una estafa.

Las estafas telefónicas no son nada nuevo, pero por lo general no es un CEO consumado.

Caso en Emiratos Árabes Unidos, 2020

- Fraude por 35 millones de dólares usando varias técnicas con las nuevas tecnologías en combinación de la clonación de voz.

Forbes AR

INNOVACION

IT y hampa: clonaron la voz de un gerente para robar US$ 35 millones a un banco

Thomas Brewster

utiliza en un gran robo en los Emiratos
vestigadores de Dubai, en medio de
la nueva tecnología por parte de los

te de un banco en los **Emiratos Árabes Unidos**
ombre cuya voz reconoció: un director de una
blado antes. El director tenía buenas noticias: su
una adquisición, por lo que necesitaba que el
ncias por valor de **35 millones de dólares**.

UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

IN RE APPLICATION OF USA PURSUANT TO 18 U.S.C. § 3512 FOR ORDER FOR COMMISSIONER'S APPOINTMENT FOR MONEY LAUNDERING INVESTIGATION

ML No: 1:21-ML-00887

Reference: DOJ Ref. # CRM-182-77215

APPLICATION OF THE UNITED STATES FOR AN ORDER
FOR A COMMISSIONER'S APPOINTMENT PURSUANT TO 18 U.S.C. § 3512

The United States of America, moving by and through its undersigned counsel, respectfully submits this *ex parte* application for an Order, pursuant to 18 U.S.C. § 3512, appointing the undersigned attorney, Rachel G. Hertz, Trial Attorney, Office of International Affairs, Criminal Division, U.S. Department of Justice (or a substitute or successor subsequently designated by the Office of International Affairs), as a commissioner to collect evidence and to take such other action as is necessary to execute this and any subsequent, supplemental requests for assistance with the above-captioned criminal matter from the United Arab Emirates (UAE). In support of this application, the United States asserts:

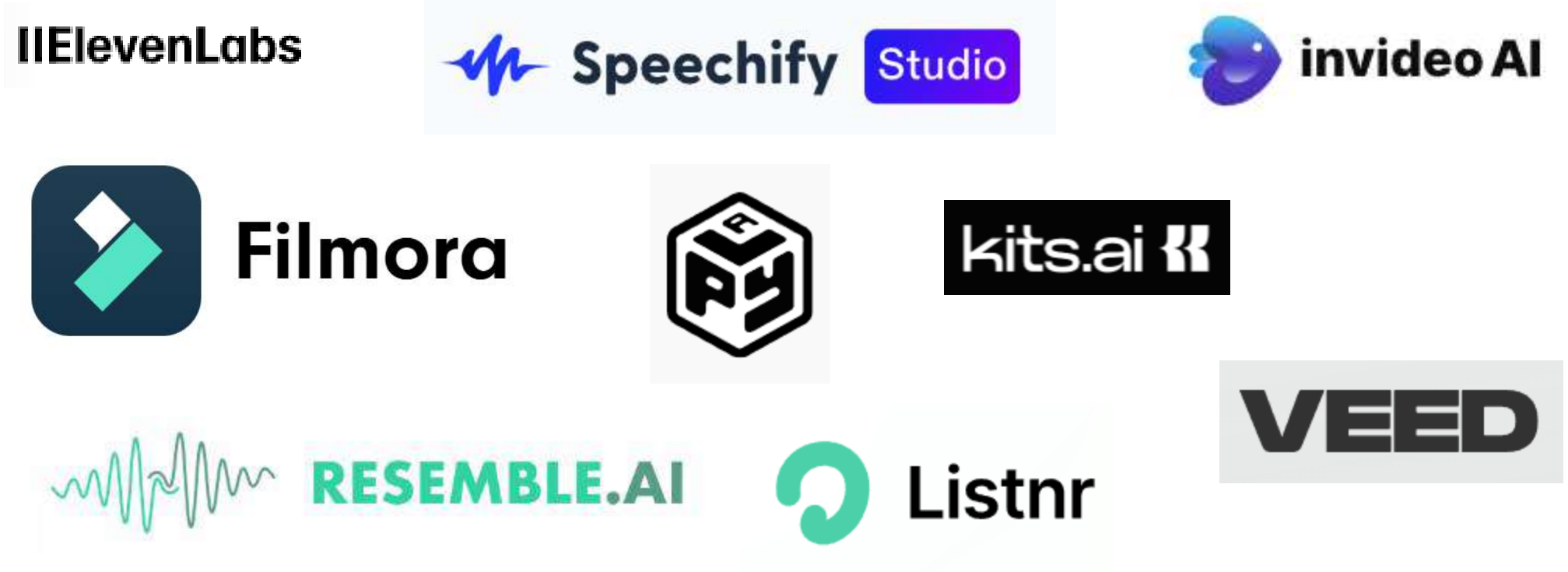
RELEVANT FACTS

1. The Ministry of Justice of the United Arab Emirates submitted a request for assistance (the Request) to the United States, pursuant to the principles of comity and reciprocity.

2. As stated in the Request, the Dubai Public Prosecution Office in the UAE is investigating multiple targets for money laundering, which occurred in or about January 2020, in violation of the criminal law of the UAE, specifically, Articles 1, 2a c, 4, 5/2-1, 1/6, 100, 21,

El crimen ha evolucionado

Decenas de herramientas gratis y pago accesibles



El crimen ha evolucionado

¿Cómo funcionan estas tecnologías?
¿Qué podemos hacer para contrarestarlas?

Características de un audio

Tono

Timbre

Velocidad de habla

Inflexiones

Real-time Detection of AI-Generated Speech for DeepFake Voice Conversion

Jordan J. Bird, Ahmad Lotfi
Nottingham Trent University
Nottingham, UK
{jordan.bird, ahmad.lotfi}@ntu.ac.uk

ABSTRACT

There are growing implications surrounding generative AI in the speech domain that enable voice cloning and real-time voice conversion from one individual to another. This technology poses a significant ethical threat and could lead to breaches of privacy and misrepresentation, thus there is an urgent need for real-time detection of AI-generated speech for DeepFake Voice Conversion. To address the above emerging issues, the DEEP-VOICE dataset is generated in this study, comprised of real human speech from eight well-known figures and their speech converted to one another using Retrieval-based Voice Conversion. Presenting as a binary classification problem of whether the speech is real or AI-generated, statistical analysis of temporal audio features through t-testing reveals that there are significantly different distributions. Hyperparameter optimisation is implemented for machine learning models to identify the source of speech. Following the training of 208 individual machine learning models over 10-fold cross validation, it is found that the Extreme Gradient Boosting model can achieve an average classification accuracy of 99.3% and can classify speech in real-time, at around 0.004 milliseconds given one second of speech. All data generated for this study is released publicly for future research on AI speech detection.

Keywords DeepFake Detection · Generative AI · Speech Recognition · Audio Signal Processing · Voice Cloning

1 Introduction

The implications of generative Artificial Intelligence (AI) in recent years are rapidly growing in importance. State-of-the-art systems capable of converting a speaker's voice to another in real-time via a microphone and sophisticated deep learning models. The ability to clone an individual's speech and use it during an online or phone call is no longer science fiction, and is possible using consumer-level computing technology.

Although this technology may prove attractive for entertainment purposes, advancements in the field pose a significant security threat. Human beings use voice as a method of recognising others in social situations and often go unquestioned. Voice recognition is also used for biometric authentication, and thus voice conversion could be used unethically to breach privacy and security. In this case, the potential for misrepresentation and identity theft are enabled, which requires immediate solutions from the scientific literature.

The scientific contributions of this work are threefold: first, the provision of an original audio classification dataset comprised of 8 well-known public figures, with real audio collected from the internet and AI-generated speech via Retrieval-based Voice Conversion (RVC). Second, the statistical analysis of extracted audio features to explore which sets of features are statistically significant given the classification of human or AI-generated speech. Third, the hyperparameter optimisation of statistical Machine Learning (ML) models towards improving accuracy and inference time, in order to achieve real-time recognition of AI-generated speech. The real-time models presented by this study are important for real-world use, and could be used, for example, to provide a warning system for individuals on phonecalls or in conference calls, where a synthetic voice may be part of the conversation with nefarious aims.

applied sciences

MDPI

Article

Detecting Deepfake Voice Using Explainable Deep Learning Techniques

Suk-Young Lim [1], Dong-Kyu Chae [1] and Sang-Chul Lee [2,*]

[1] Department of Artificial Intelligence, Hanyang University, Seoul 04763, Korea; offlim@hanyang.ac.kr (S.-Y.L.); dongkyu@hanyang.ac.kr (D.-K.C.)
[2] Division of Nanotechnology, Daegu Gyeongbuk Institute of Science & Technology (DGIST), Deagu 42988, Korea
* Correspondence: sangchul.lee@dgist.ac.kr

Abstract: Fake media, generated by methods such as deepfakes, have become indistingu
from real media, but their detection has not improved at the same pace. Furthermore, the
of interpretability on deepfake detection models makes their reliability questionable. In thi
we present a human perception level of interpretability for deepfake audio detection. Ba
their characteristics, we implement several explainable artificial intelligence (XAI) metho
for image classification on an audio-related task. In addition, by examining the human c
process of XAI on image classification, we suggest the use of a corresponding data for
providing interpretability. Using this novel concept, a fresh interpretation using attributio
can be provided.

Keywords: explainable artificial intelligence (XAI); deepfake detection; human-centered a
intelligence

check for updates

Citation: Lim, S.-Y.; Chae, D.-K.; Lee, S.-C. Detecting Deepfake Voice Using Explainable Deep Learning Techniques. *Appl. Sci.* 2022, 12, 3926. https://doi.org/10.3390/app12083926

Academic Editors: Andrea Prati, Luis Javier García Villalba and Vincent A. Cicirello

Received: 28 February 2022
Accepted: 11 April 2022
Published: 13 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

With recent advances in artificial intelligence and its applications, cases of ab
AI technology have also increased. A deepfake is one of the main methods that
many of such cases. Thus far, only a few celebrities have been targeted. However, o
two phenomena triggered by the public's recent increased use of social media, i.e.,
of data collection and (2) enhanced influence of information distribution, fake me
proliferated.

While deepfake generation has improved considerably in recent times, the ac
of deepfake detection has remained at 82.56% when tested upon a public open dat
Though this performance improvement is significant from an academic perspective, i
insufficient for real-world usage. Given two major emerging issues, i.e., less-than-
accuracy of detection and widened target range, interpretability of deepfake detect
become a critical consideration. However, contemporary research on explainable de
detection is not extensive and is limited to visual deepfake detection [2].

In this study, we implemented XAI methods on deepfake voice detection in o
be able to recommend the proper delivery of the interpretation at a human perceptio
To target the non-experts for linguistics as well as artificial intelligence, the study is fo
on intuitiveness and a higher level of interpretability.

Currently, for speech recognition or speaker verification, methods, such as transform-
ers, conformers, or wav2vec, already show good performance [3–5]. However, in this study,
to focus on the proper delivery of the interpretation rather than the performance, simple

Research Square

Deepfake audio detection and justification with Explainable Artificial Intelligence (XAI)

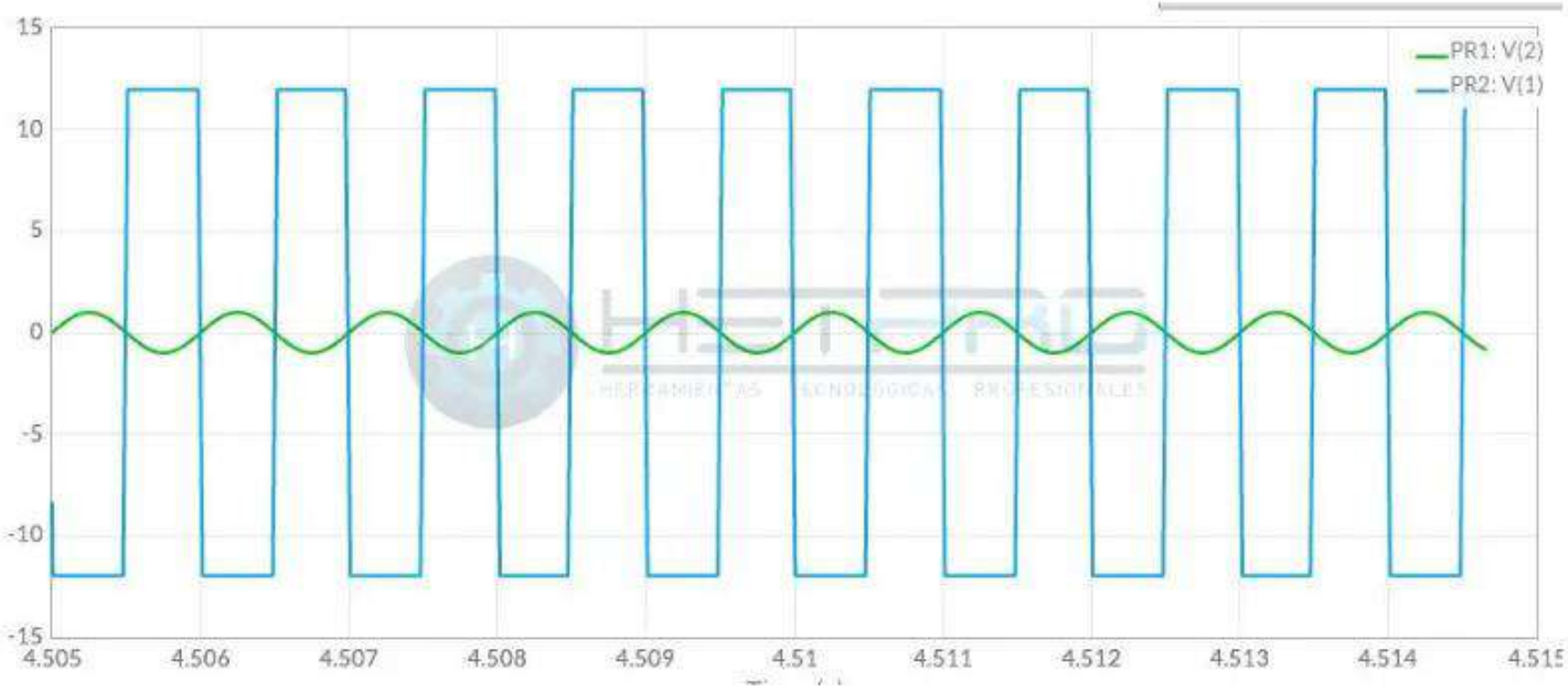
Aditi Govindu (adigovindu@gmail.com)
MIT World Peace University
Preeti Kale
MIT World Peace University
Aamir Hullur
MIT World Peace University
Atharva Gurav
MIT World Peace University
Parth Godse
MIT World Peace University

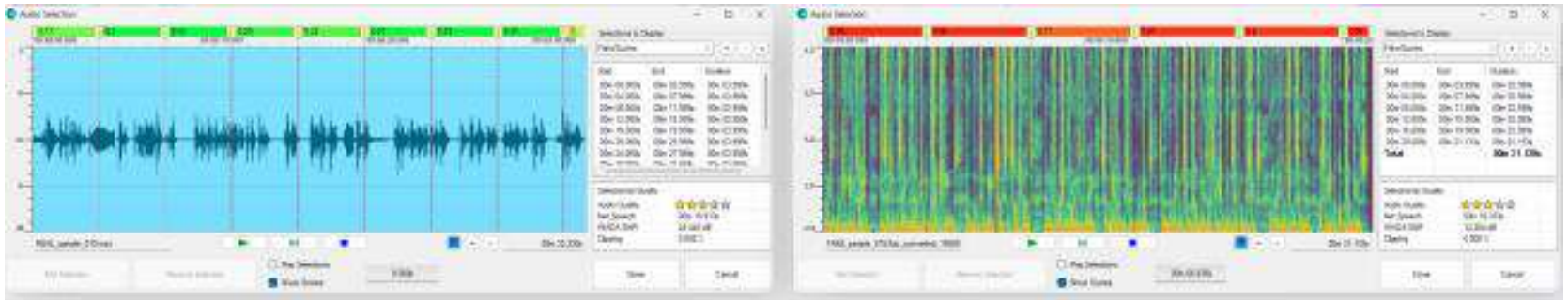
Research Article

Keywords: Generative Adversarial Neural Networks (GANs), deepfake audio, VGG16, Explainable Artificial Intelligence (XAI), Fréchet Audio Distance (FAD)

Posted Date: October 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3444277/v1


15
10
5
0
-5
-10
-15
4.505
4.506
4.507
4.508
4.509
4.51
4.511
4.512
4.513
4.514
4.515
PR1: V(2)
PR2: V(1)



Combatiendo los DeepFake de video

El secto privado apoyando a las autoridades francesas

Un desafío a largo plazo

Adopción de la tecnología por los criminales

Gracias por su atención