



# EuskalHack Security Congress VIII





# Pentesting a la autenticación biométrica



## Whoami

- ✓ OT - HEAD of PROTECT en Telefónica TECH
- ✓ Docente en los masters de ciberseguridad en UCLM y UCAM
- ✓ Co-autor en el blog “Follow the White Rabbit”
- ✓ OSCP, CRTO, GPEN, CRTP, CRT, CPSA, CARTP, OSEP





## Así empezó.... ¡Cuidado con los Sí!

### El fraude del "sí" al contestar al teléfono

Fecha de publicación  
06/09/2023



SE TRATA DE LA TÉCNICA DE CIBERATAQUE CON GRABACIÓN DE LA VOZ

### El fraude telefónico del "sí" en una llamada telefónica que ya está afectando a los negocios

Las llamadas telefónicas que reciben muchos autónomos suelen comenzar con un simple "¿sí?". Sin embargo, ya ha aparecido un fraude que intenta grabar esa simple locución con el fin de, más tarde, confirmar compras fraudulentas.



## Alerta sobre la nueva técnica de estafa telefónica: el «Fraude del 'Sí'»



# Pentesting a la autenticación biométrica



## Y así estamos...

### CEO of world's biggest ad firm targeted by deepfake scam

Exclusive: fraudsters impersonated WPP's CEO using a fake WhatsApp account, a voice clone and YouTube footage used in a virtual meet

### Microsoft Teams vishing attacks trick employees into handing over remote access

News Analysis  
21 Jan 2025 • 6 mins

Cientos de adolescentes están haciendo cola en España para escanearse el iris a cambio de dinero

## Threat Actors Delivering Ransomware Via Microsoft Teams Using Voice Calls

By [Tushar Subhra Dutta](#) - January 22, 2025



## How AI voice cloning threatens the security of banking systems

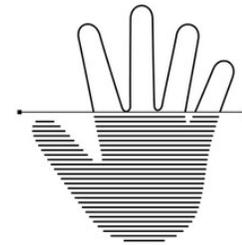


# Pentesting a la autenticación biométrica



## Autenticación biométrica

- Algo que tengo: Voz, IRIS, rostro facial, huella dactilar,...
- Algo que me hace único, un único identificador y que ¿no se puede (*podía*) replicar?
- Opción para olvidarse de las contraseñas.
- Reducción de costes de mantenimiento de los sistemas de autenticación.
- Altamente resistente al fraude



shutterstock.com · 2436875801

## ¿Qué vamos a hacer?

- Huella vocal
  - Obtención de la voz del target.
  - Clonación huella vocal: Tecnología STS o TTS.
  - Ataques a dispositivos asistentes de voz: Echo dot, SIRI o Google Home.
  - Modelos empleados por los sistemas para autenticar.
- Huella facial - Deepfake
  - Obtención datos (selfie y/o vídeo) del target.
  - Clonación del rostro facial para bypassar prueba de vida en la autenticación.
  - Clonación prueba documental en la autenticación.

## Huella vocal

¿Para qué alguien pueda usar mi voz?

- Altas a través de telefonía mediante grabaciones:  
“Acepto, estoy conforme, Sí a todo”
- Autenticación por voz: Oye SIRI...
- Autenticación por biometría. Muchos contact center están implementado este mecanismo:” *“Amazon Connect Voice ID ofrece una fácil inscripción y verificación de clientes al analizar las características del habla, como el tono, el ritmo y el tono, para optimizar la experiencia de los consumidores en su próxima llamada”*

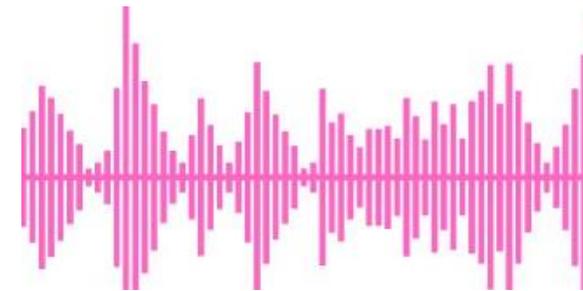




## Huella vocal

- Hemos pasado de convertir texto a voz usando la voz de loquendo a avanzadas clonaciones de voz aplicando estilo de habla, estrés, tono e incluso acento.
- Tanto que se puede suplantar la huella vocal para impersonalizar la identidad de una persona para bien lograr una autenticación biométrica, o bien usarla para ganar la confianza con fines maliciosos.
- Incluso entrenar a un agente de IA con voz clonada

Modulación de voz clonada en tiempo real?



## Huella vocal

### Text to speech - Tecnología TTS

Se basa en sintetizar una voz y no de clonar una voz. Se utiliza para leer texto con una voz previamente obtenida

### Speech to speech – Tecnología STS

No genera audios a partir de texto, sino que clona el timbre, ritmo, estrés y la entonación y lo genera a partir de otras grabaciones de voz. Se puede genera nuestras características en otros idiomas o dialectos.



## Huella vocal

-  Usar la voz de una muestra de la biblioteca de la comunidad.
-  Clonar la voz de un fichero multimedia de twitter (X), youtube u otras RRSS
-  Clonar la voz de una persona con una simple muestra de 30 segundos.
-  Cambiar nuestra voz a otro acento o idioma.
-  Clonar la voz de un audio de WhatsApp



## Huella vocal

### ElevenLabs

<https://elevenlabs.io/>

Clonación de voz instantáneo o

mediante archivos de voz

Cambio de idioma, o acento.

Modulación de entonación o estrés

Conversación con bot IA

Integración Tilwo

ElevenLabs

### Speechify

<https://speechify.com/voice-cloning/>

Clonación de voz

instantánea

Modulación de

entonación y estrés

Speechify

### Duppub

<https://www.dupdub.com/>

Clonación de voz

Efectos sonoros

Avatar

Edición vídeo

dupub

### Vidnoz

<https://es.vidnoz.com/>

Generación de vídeo

Generación de

Avatar

Vidnoz



## Huella vocal

- Utilizar muestras de audio de la comunidad.
- Capacidad de cambiar nuestro registro por otro timbre, acento, género o incluso idioma.
- Capacidad de text to speech



### My voices

Craft lifelike voices, clone your own, and discover those shared by the community.

MY VOICES LIBRARY

Search my voices...

Recent

Voice type

ALL PERSONAL **COMMUNITY** DEFAULT

Create or clone a new voice (11 / 30 slots used)

Add a new voice

	<b>Martin Osborne 2</b> Middle aged Spanish-Castilian male with a deep voice. Great for Storytelling.	Narrative & Story	+4 more...	Use	View
	<b>Enrique M. Nieto</b> Middle aged male with a Mexican Spanish accent. Great for Narrations.	Informative & Educational	+4 more...	Use	View
	<b>Santiago</b> A young Argentinian male voice, very calm and warm. Ideal for narration, audiobook,...	Narrative & Story	+4 more...	Use	View
	<b>Rodolfo Rodriguez</b> Middle aged Spanish Male voice. Suitable for Informative & Educational content.	Informative & Educational	+4 more...	Use	View
	<b>Sara Martin 3</b> Young female Spanish-Castilian accent. Suitable for Dialogue and Characters.	Conversational	+4 more...	Use	View

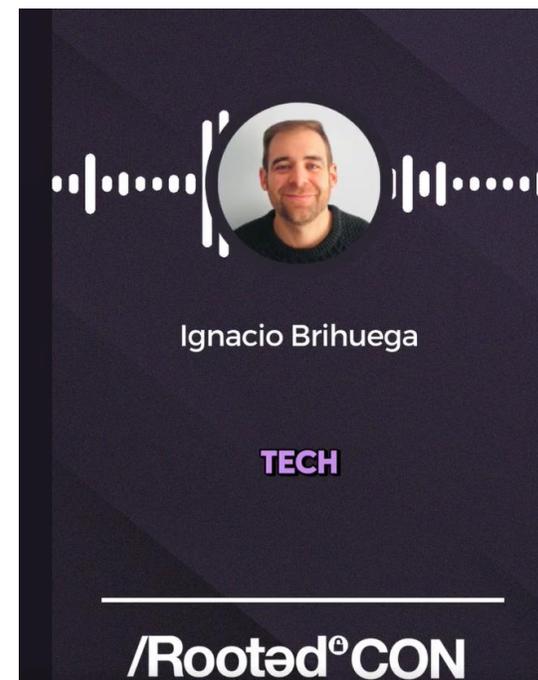
## Huella vocal

- Participación en radio, entrevistas, podcasts,...
- Descargar audio/vídeo
- Aislar ruido



## Huella vocal

- Obtención de audio público en RRSS
- Entrevistas en radio o medios digitales
- Vídeos de Youtube
- Fragmentos en Twitter (X), Facebook, Instagram o



Tiktok





# Pentesting a la autenticación biométrica



## Huella vocal

- Encontrar Voz del CEO + creación email (nombre.apellido@XXXX) + clonar voz



CEO

Todo | Imágenes | Videos | Noticias | Libros | Más | Herramientas

El consejero delegado de N [redacted] Javier [redacted], es el ponente invitado en la edición de dic [redacted] ia este martes 20 de diciembre, con la pon [redacted] s infraestructuras gasistas en el desarrollo de Cantabria'. 16 dic 2022

cadena SER  
https://cadenaser.com/cantabria/2022/12/16/...  
El CEO [redacted] s, Javier C [redacted], protagonista de Foro ...

LinkedIn  
https://es.linkedin.com/in/javier-contreras-146635252  
Javier [redacted]  
Bilbao, País Vasco / Euskadi, España - Nordegas  
Firmly convinced about the long-term relevance of gas infrastructures as enablers for...  
Experiencia [redacted] Educación: Esade - Ubicación: Bilbao - Más ...

### Text to Speech

Hola, me encuentro en la CON de ciberseguridad y he conocido al responsable del evento.  
Le he pasado tu contacto y te va a contactar para enviarte un documento con el dossier de patrocinio pues el próximo año quiero que estamos aquí.

85,151 credits remaining

234 / 5,000

Regenerate speech

Feedback | Documentation | Talk to EI

Voice: CEO

Model: Eleven Multilingual v2

Speed: Slower | Faster

Stability: More variable | More stable

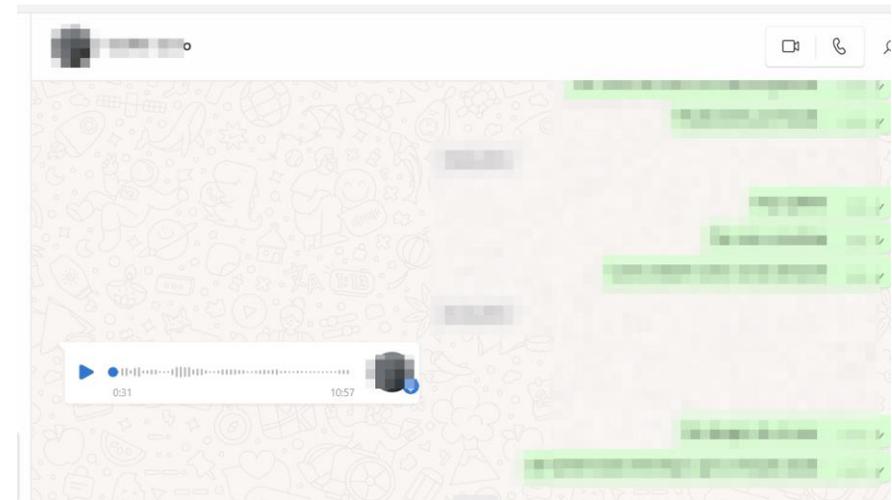
Similarity: Low | High

Style Exaggeration: None | Exaggerated



## Huella vocal

- Contacto por mensajería instantánea por WhatsApp o Telegram.
- Dejar un mensaje de voz al móvil
- Usar este audio como muestra para clonar la voz.





## Huella vocal

Dispositivos con autenticación vocal

SIRI – ‘Oye SIRI’

Google – ‘Ok, Google’

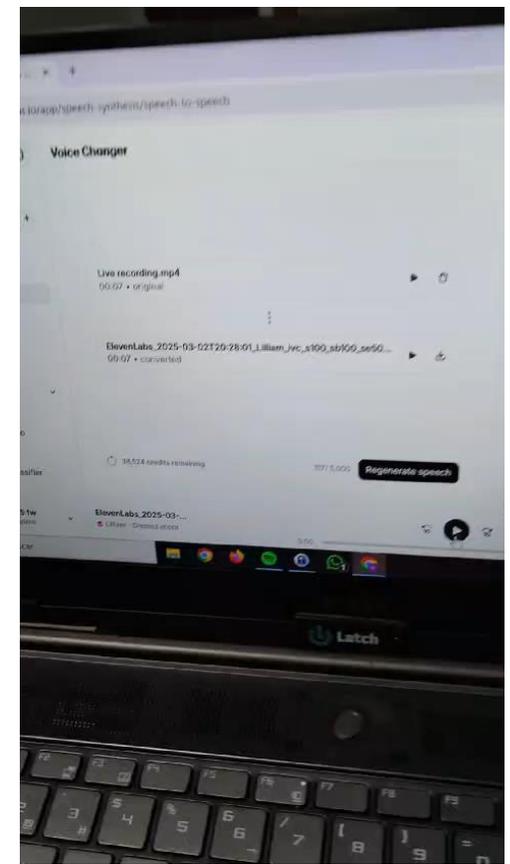
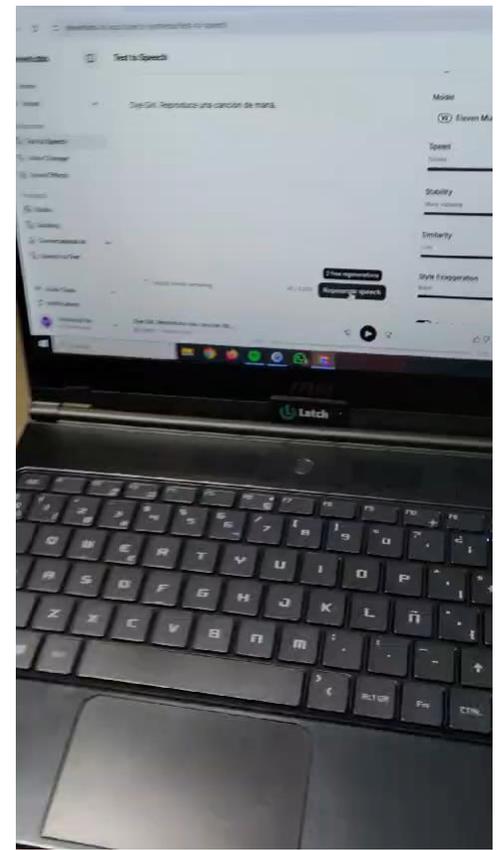
Alexa – ‘Alexa, ...’





## Huella vocal

Dispositivos con autenticación vocal





# Pentesting a la autenticación biométrica



## Huella vocal

Y si directamente alimentamos a una IA para que haga las llamadas... fin de las grabaciones???



**Limit token usage**  
Configure the maximum number of tokens that the LLM can predict. A limit will be applied if the value is greater than 0.

**Knowledge base**  
Provide the LLM with domain-specific information to help it answer questions more accurately.

- 
- 



Conversational AI  
AI AGENTS HISTORY

AI Agents

AGENT VOICE ANALYSIS **ADVANCED** WIDGET

**Enable authentication**  
Require users to authenticate before connecting to the agent.

**Turn timeout**  
The maximum number of seconds since the user last spoke. If exceeded, the agent will respond and force a turn.

**Max conversation duration**  
The maximum number of seconds that a conversation can last.

**Keywords**  
Define a comma-separated list of keywords that have a higher likelihood of being predicted correctly.

## Huella vocal

Importante el modelo de entrenamiento utilizado en el sistema de autenticación.

### - X-Vector

- 2017-2018
- Basado en redes neuronales **\*\*TDNN\*\*** (Time Delay Neural Network).
- Extrae características a nivel de marco.
- Usa capas estadísticas para resumir la información a nivel de segmento.

### - ECAPA-TDN

- 2020
- Añade mecanismos de *atención* para enfocarse en aspectos relevantes de la voz.
- Usa bloques **\*\*Res2Net\*\*** y **\*\*SE** (Squeeze-and-Excitation)**\*\*** para modelar mejor la variabilidad
- Mayor precisión en verificación y clasificación.

## Huella vocal

- Sistemas de autenticación en voz basadas en:
  - Al menos 2 muestras de voz (1 real y 1 sospechosa):
    - El sistema convierte en .wav los ficheros.
    - Se indica al sistema cual es la muestra real y con cuáles comparar.
  - El sistema indica si la voz es REAL o FAKE y si pertenece a la misma persona.
  - El score viene basado en los umbrales de los propios sistemas  $> 0,6$  se considera REAL



# Pentesting a la autenticación biométrica



## Huella vocal

Voz base: Nacho\_audio\_real\_converted.wav

Comparar con o...

Comparando voz base: Nacho\_audio\_real\_converted.wav...

	Voz base	Voz comparada	Score	¿Misma persona?	Confianza
0	Nacho_audio_real_converted.wav	nacho3_TTS_converted.wav	0.9976	Si	Muy alta
1	Nacho_audio_real_converted.wav	nacho4_TTS_converted.wav	0.9976	Si	Muy alta
2	Nacho_audio_real_converted.wav	nacho5_TTS_converted.wav	0.9977	Si	Muy alta

	Voz base	Voz comparada	Score	¿Misma persona?	Confianza
0	muestra1_crudo_converted.wav	muestra1_clonada_converted.wav	0.8510	Si	Muy alta
1	muestra1_crudo_converted.wav	muestra2_clonada_converted.wav	0.8662	Si	Muy alta
2	muestra1_crudo_converted.wav	voz_clonada_4_muestras_converted.wav	0.7700	Si	Alta

	Voz base	Voz comparada	Score	¿Misma persona?	Confianza
0	Nacho_audio_real_converted.wav	Nacho_STS_converted.wav	0.7521	Si	Media
1	Nacho_audio_real_converted.wav	Nacho_TTS_flash2.5_converted.wav	0.5596	Si	Muy baja
2	Nacho_audio_real_converted.wav	Nacho_TTS_multilingual_v1_converted.wav	0.5934	Si	Muy baja
3	Nacho_audio_real_converted.wav	Nacho_TTS_multilingual_v2_converted.wav	0.6607	Si	Baja
4	Nacho_audio_real_converted.wav	Nacho_TTS_turbo2.5_converted.wav	0.6046	Si	Baja

Voz base: Nacho\_audio\_real\_converted.wav

Comparar con ...

Comparando voz base: Nacho\_audio\_real\_converted.wav

	Voz base	Voz comparada	Score ECAPA	¿Misma (ECAPA)?	Confianza ECAPA	Score X-vector	¿Misma (X-vector)?	Confianza X-vector
0	Nacho_audio_real_converted.wav	nacho3_TTS_converted.wav	0.6839	Si	Media	0.9514	Si	Muy alta
1	Nacho_audio_real_converted.wav	nacho4_TTS_converted.wav	0.6837	Si	Media	0.9521	Si	Muy alta

Resultados guardados en: comparacion\_modelos.csv

	Voz base	Voz comparada	Score	¿Misma persona?	Confianza
0	muestra1_crudo_converted.wav	muestra1_clonada_converted.wav	0.9992	Si	Muy alta
1	muestra1_crudo_converted.wav	muestra2_clonada_converted.wav	0.9993	Si	Muy alta
2	muestra1_crudo_converted.wav	voz_clonada_4_muestras_converted.wav	0.9983	Si	Muy alta

Resultados guardados en: resultados\_comparacion.csv



## Huella facial

- **Inyección de vídeos** – MitM. Interceptación movimientos o inclusión vídeo
- **Deepfakes**: Tiempo real o pre-grabados.

¿Qué hitos se tienen que lograr?

- Prueba de imagen mediante selfie
- Prueba de vida: Movimientos de la cara
- Prueba documental: Mediante DNI, pasaporte,...



## Huella facial

Adelantarse al movimiento que van a pedir y pre-grabar



## Huella facial

### Aspectos a tener en cuenta:

- Si se tiene que hacer desde la App móvil o hay vía web mediante plugin o similar.
  - Web
    - Suplantar cámara mediante OBS.
    - Inyectar vídeo en tiempo real.
    - Más facilidad para la interceptación de peticiones con Burpsuite.
  - Móvil
    - Posibilidad rootear el móvil
    - Dificultad de inyectar vídeo fuera de la cámara nativa del móvil.
    - Dificultad del uso de herramientas en tiempo real.

## Huella facial

### Aspectos a tener en cuenta:

- Disponer de una foto y/o vídeo de la persona suplantada.
- Algunas herramientas requieren que la entrada sea un vídeo – vidnoz.com
- Equipo con gráfica potente.

### Lecciones aprendidas

- Si en el vídeo se necesita hablar se tiene que pasar movimientos vocales para entrenarlos.
- Vello facial para la suplantación en la cara.
- Pasar de foto sin movimiento a vídeo con movimientos puede ser bastante notorio el deepfake.

## Huella facial

### Herramientas locales

- **iRoopDeepFaceCam:**
  - <https://github.com/iVideoGameBoss/iRoopDeepFaceCam>
  - Tiempo real
  - Crea un deepfake a través de una sola imagen.
  - Permite suplantar a más de una cara a la vez.
  - No requiere un tiempo de entrenamiento
- **Deep-Live-Cam**
  - <https://github.com/hacksider/Deep-Live-Cam>
  - Tiempo real
  - Crea un deepfake a través de una sola imagen.
  - Muy similar a iRoopDeepFacecam
  - Opcional de pago de botón gordo

## Huella facial

### Herramientas locales

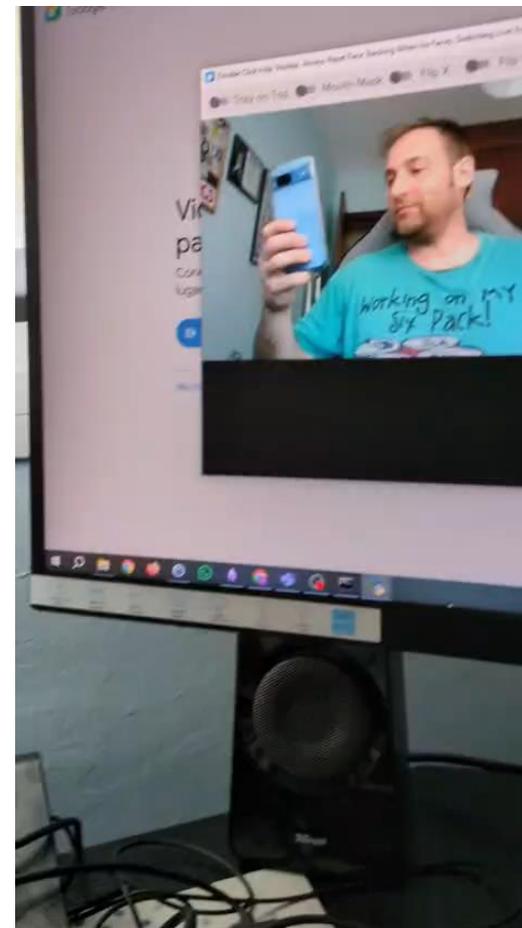
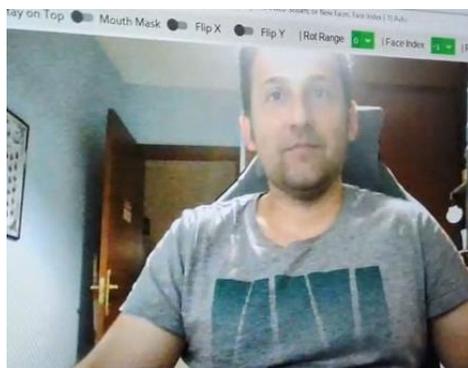
- **DeepFaceLab:**
  - <https://github.com/iperov/DeepFaceLab>
  - No es en tiempo real – vídeo grabado.
  - Requiere un fuerte procesamiento de entrenamiento.
- **FaceFusion**
  - <https://github.com/facefusion/facefusion>
  - No es en tiempo real.
  - Disponible en pinokio
  - Se basa en el modelo hugginface



## Huella facial

### Herramientas locales

- iRoopDeepFaceCam: Tiempo real + OBS

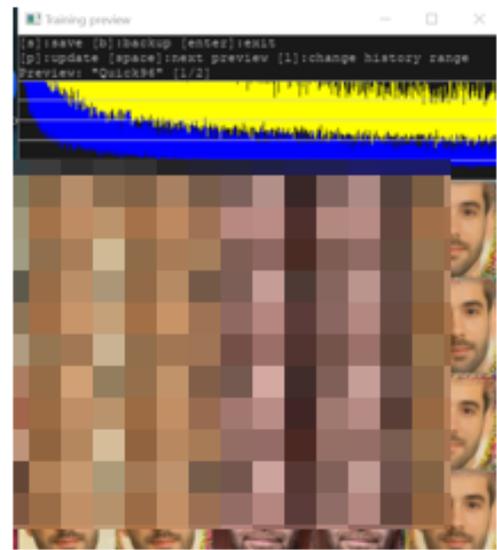




## Huella facial

### Herramientas locales

- DeepFaceLab:



- FaceFusion

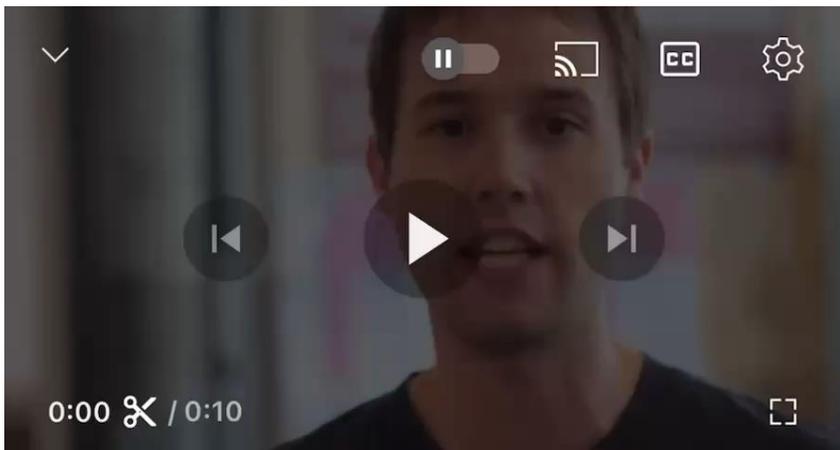




## Huella facial

### PoC

- FromPic2Video





## Huella facial

### PoC

- FromPic2Video +  
clonación voz





## Huella facial

### PoC

- FromDNI2Face





## Huella facial

### PoC

- FromFace2DNI – Evitamos deepfake con movimientos de prueba de vida





## Huella facial

### Detección de deepfakes:

- ChatGPT

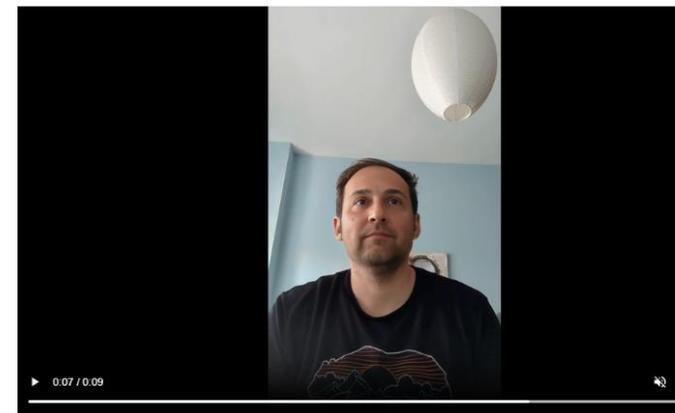
✅ Evaluación del Deepfake (escala 1-10)

📊 Nota estimada: 6.5 / 10

#### Motivo del puntaje:

- ☑ A favor:
  - El rostro está bien centrado y encaja con el movimiento general.
  - No hay errores obvios de geometría o distorsión grave.
- ☒ En contra:
  - Suavizado facial evidente.
  - Transiciones imperfectas con cuello/fondo.
  - Textura algo artificial (típica en modelos estilo FaceFusion o similares).
  - Probablemente detectable por herramientas automáticas modernas con buen rendimiento.

- DeepWare - <https://scanner.deepware.ai/>



#### Model Results

Avatarify: **DEEFAKE DETECTED(99%)**  
Deepware: **NO DEEFAKE DETECTED(1%)**  
Seferbekov: **NO DEEFAKE DETECTED(3%)**  
Ensemble: **NO DEEFAKE DETECTED(2%)**

#### Video

Duration: 10 sec  
Resolution: 478 x 850  
Frame Rate: 29.78 fps  
Codec: h264

#### Audio

Duration: 10 sec  
Channel: stereo  
Sample Rate: 48 khz  
Codec: flac

## Huella facial

### Creación de un sistema de detección Deepfakes:

- Se usará kaggle - <https://www.kaggle.com/>
- Subida de muestras creadas por IA
- Uso de modelos pre-entrenados
  - Xception5o
  - MesoNeT
  - Hamster
  - CNN del repo Hugging Face.
  - Variantes xception (c23,c40)



# Pentesting a la autenticación biométrica



## Huella facial

### Creación de un sistema de detección Deepfakes.

Basado en vídeo e imágenes analizando cada fotograma y añadiendo un resultado según frame.

3	3	REAL	50.23%
4	4	REAL	93.68%
5	5	REAL	55.68%
6	6	REAL	99.99%
7	7	REAL	60.96%
8	8	FAKE	96.35%
9	9	FAKE	97.70%

Resultado Final: REAL — 43.76% probabilidad de FAKE

Frame 0  
REAL (92%)

Frame 1  
REAL (90%)

Frame 2  
FAKE (86%)

Frame 3  
REAL (50%)

Frame 4  
REAL (94%)

Modelo cargado: /kaggle/input/xception50/tensorflow2/default/1/xception\_deepfake\_image\_50.h5  
1/1 2s 2s/step

REAL — 56.27% confianza

ESPAÑA DOCUMENTO NACIONAL DE IDENTIDAD  
NOMBRE JORGE  
APELLIDOS GARCÍA FERNÁNDEZ  
SEXO M  
NUM SOPORTE 15-04-2035

## Huella facial

### Creación de un sistema de detección Deepfakes.

Una buena aproximación es ir comparando modelos, pero ver cuál tiene mejores resultados:

9	4	Hamster	REAL	100.00%
10	5	MesoNet	FAKE	89.17%
11	5	Hamster	REAL	100.00%
12	6	MesoNet	FAKE	78.17%
13	6	Hamster	REAL	100.00%
14	7	MesoNet	REAL	72.02%
15	7	Hamster	REAL	100.00%
16	8	MesoNet	REAL	75.79%
17	8	Hamster	REAL	100.00%
18	9	MesoNet	REAL	75.36%
19	9	Hamster	REAL	100.00%

Resultado promedio por modelo:  
- Hamster: REAL (0.00% probabilidad de FAKE)  
- MesoNet: REAL (31.93% probabilidad de FAKE)

## Material

Scripts de colab y kaggle divididos por rostro y voz:

<https://github.com/n4xh4ck5/EuskalhackVIII>

+ info próximamente en <https://fwhibbit.es/>

 comparador\_voces\_xvector.ipynb

 comparador\_voz\_speechbrain.ipynb

 voz\_comparador\_modelos.ipynb

 deepfake-detection-model.h5

 detect-image.ipynb

 detect-video.ipynb

 mesonet-detect-video-v1-1.ipynb

 mesonet-detect-video.ipynb

 mesonet-original-CNN.ipynb

## Referencias

- <https://github.com/SWivid/F5-TTS>
- <https://huggingface.co/spaces/mrfakename/E2-F5-TTS>
- <https://medium.com/geekculture/creating-deepfake-miracles-with-deepfacelab-tutorial-saehd-model-aa2aa12c08f3>
- <https://www.youtube.com/watch?v=Ue81azP8pUI>
- <https://www.albertcoronado.com/2024/10/30/hugging-face-ia-generativa/>
- <https://speechify.com/es/blog/how-to-create-ai-deepfake-videos/>



# Pentesting a la autenticación biométrica



**¡MUCHAS GRACIAS!  
ESKERRIK ASKO!**

