

UNIVERSIDAD COMPLUTENSE DE MADRID

**FACULTAD DE ESTUDIOS ESTADÍSTICOS**

**MÁSTER EN BIG DATA, DATA SCIENCE & AI**

TRABAJO FINAL DE MÁSTER

**Detección de DeepFake Voice con análisis de contenido fraudulento mediante un enfoque multimodal de Deep Learning**

Convocatoria de septiembre 2025

[Aquí poner una frase celebre o de algo que pueda representar al TFM]

¿Algo así?

"El espíritu humano debe prevalecer sobre la tecnología."

- Albert Einstein, físico.

Agradecimientos

[Poner agradecimientos, no es obligatorio – Este apartado es opcional]

Resumen

Resumen del proyecto a grandes rasgos, que ocupe aproximadamente mitad de hoja

Palabras claves

Palabras claves (pondría máximo 8, eso como lo veáis)

Abstract

SAME AS BEFORE BUT IN ENGLISH

Keywords

SAME AS BEFORE BUT IN ENGLISH

Índice de Contenidos

[1. Introducción 13](#_Toc206593904)

[2. Investigación previa 17](#_Toc206593905)

[3. Objetivos 19](#_Toc206593906)

[Objetivos específicos 19](#_Toc206593907)

[4. Desarrollo de la solución técnica 20](#_Toc206593908)

[5. Resultados 22](#_Toc206593909)

[6. Implicaciones éticas e impacto social 24](#_Toc206593910)

[7. Conclusiones 26](#_Toc206593911)

[8. Bibliografía 28](#_Toc206593912)

[Anexos 29](#_Toc206593913)

Índice de Tablas

Índice de tablas

Índice de Figuras

Índice de figuras/imágenes

Lista de Acrónimos

|  |  |
| --- | --- |
| **Acrónimo** | **Significado** |
| IA | Inteligencia Artificial |
| DL | Deep Learning |
| SMS | *Short Message Service* |

IMPORTANTE: Ordenados en orden alfabético

# Introducción

Los casos de estafa telefónica cada vez son más comunes. Consiste en suplantar la identidad y datos biométricos como la voz de personas cercanas o conocidas de la víctima, para obtener recompensas económicas a cambio. Así pues, simulando una situación de emergencia, debido a la rapidez de los hechos, la víctima puede realizar las acciones solicitadas sin cuestionar su legitimidad. La manipulación emocional, al escuchar la voz conocida, bloquea el juicio crítico hasta el punto de efectuar transferencias, compartir códigos de seguridad o divulgar información personal sensible.

Así fue el caso de *Arup*, una multinacional británica de diseño e ingeniería, que, tras unas videollamadas con datos biométricos replicadas de compañeros y superiores, un trabajador de finanzas de Hong Kong transfirió un total de 25,6 millones de dólares a los estafadores. Este fraude fue más allá, pues no sólo suplantaron las voces si no la imagen de sus compañeros, siendo todos los participantes de la reunión recreaciones falsas menos la víctima. <https://es.tradingview.com/news/invezz:d3919989709cd:0/>

Pero no hace falta pensar en grandes empresas y millones de dólares para darse cuenta de la gravedad del problema, pues cada vez son más los casos en los que personas vulnerables, ciudadanos de a pie, sufren pérdidas que, aunque no sumen millones, son devastadores a nivel personal. El problema reside en que el oído humano no está entrenado para detectar esos sutiles artefactos digitales que delatan las voces sintéticas.

Existe una asimetría fundamental, pues para un sistema de IA, una voz no es más que una colección de señales digitales detectables pudiendo identificar una inconsistencia estadística, mientras que, para la confianza humana, es una simple llamada de un conocido pidiendo ayuda.

El gobierno de España ha tomado medidas sobre esta situación, pues con el plan anti-estafa han bloqueado 48 millones de llamadas y 2,2 millones de SMS fraudulentos. Entró en vigor en marzo de este año bajo el ministro de Transformación Digital, Óscar López. De la mano de las operadoras, la medida principal de este plan consiste en usar un nuevo prefijo específico de tres cifras para toda llamada que se realice con fines comerciales.

Otras de las medidas planificadas para 2026 consta de crear una base de datos oficial, que incluya los códigos alfanuméricos e identificadores de las empresas y administraciones. De esta manera, se podrá identificar y corroborar las llamadas y mensajes recibidos con el fin de reducir considerablemente los fraudes.

<https://elderecho.com/bloqueados-millones-de-llamadas-y-de-sms-fraudulentos-con-el-plan-antiestafas-del-gobierno>

<https://www.eldiario.es/tecnologia/lopez-asegura-han-bloqueado-48-millones-llamadas-fraudulentas-admite-agujero-extranjero_1_12541143.html>

También se ha pronunciado José García Serrano, Comisario Principal de la Policía Nacional, responsable de la Unidad Central de Ciberdelincuencia, afirmando que el 80% de las llamadas que imitan ser de un banco para cometer una estafa, ya usan la Inteligencia Artificial. Reitera la importancia de concienciar a la población ante las crecientes amenazas derivadas de la ciberdelincuencia, y es que, en 5 años, se ha multiplicado por dos y medio la cantidad de denuncias asociadas a ciberdelitos, siendo cada vez más sofisticadas y convincentes. <https://www.atresmedia.com/metafuturo/garcia-serrano-comisario-ciberdelincuencia-cinco-anos-multiplicado-dos-medio-numero-denuncias-relacionadas-ciberdelitos_20221121637b7b622d34b10001c4ba33.html>

La verdadera frontera del fraude telefónico ya no sólo reside en el *telemarketing* masivo, sino que, debido a los avances tecnológicos, ya adquieren ataques de ingeniería social dirigidos. Es en este punto donde las regulaciones actuales se encuentran limitadas, y donde este proyecto demuestra su vital importancia.

Este no se enfoca en el ruido de las llamadas comerciales masivas, sino en el ataque dirigido que clona la voz de personas conocidas por la víctima. Para ello, es importante comprender dos puntos de estas llamadas:

* La voz del locutor: La que es generada por IA
* La transcripción: A partir de eso, se podría determinar si la llamada es fraudulenta. Por ejemplo, si el locutor está pidiendo dinero, una transferencia rápida, o códigos de verificación, esto aumenta la confianza del modelo para determinar que se trata de una llamada *scam*.

A lo largo de este documento, se expone un análisis exhaustivo del estado del arte, indagando tanto en tecnologías de generación de voz sintéticas utilizadas en dichos fraudes, como proyectos y librerías de Python existentes que ayuden a su detección. A posteriori, se profundiza en las metodologías y arquitecturas propuestas, diferenciando las distintas capas de análisis consideradas y su desarrollo correspondiente.

Con el fin de productivizar la aplicación, se incluye un apartado de implementación y despliegue, en el que se detalla cómo se empaqueta la aplicación para garantizar su uso y aplicabilidad optimizando la experiencia del usuario. El resultado final se presenta como un ejecutable orientado a todo tipo de público sin necesidad de conocimientos técnicos, que simplemente disponga de un ordenador al que pueda enlazar el teléfono móvil por el que vaya a recibir las llamadas.

Finalmente, se exponen y analizan los resultados obtenidos en las pruebas realizadas. A continuación, se expone un debate sobre el marco ético-legal, incluyendo las normativas gubernamentales y europeas a las que esta aplicación se debe ceñir. Se concluye con un capítulo en el que se recapitulan los hallazgos clave de este proyecto, además de evaluar el grado de cumplimiento de los objetivos expuestos, y futuras líneas de investigación para expandir los límites de la aplicación.

NO SÉ SI INCLUIRLO EN INTRO O DÓNDE

Diagrama

El contenido generado por IA puede ser incorrecto.La lógica es la siguiente: si el modelo detecta que la voz de la persona que llama es generada por IA, automáticamente lo clasifica como voz sintética, alarmando al usuario de que se trata de una llamada sospechosa. Sin embargo, si no se da este caso, es decir, lo detecta como voz humana, se procede a extraer la transcripción para determinar, mediante un modelo de clasificación si se trata de *scam*. Es decir, la lógica del programa es la siguiente:

# Investigación previa

Aquí ya desarrollar un poco más la teoría acerca de lo que trata el TFM en este caso modelos multimodales, audios, etc…

Estado del arte – Que hay ahora mismo en el mercado, que es lo que ya se ha hecho

Que cosas puede aportar este proyecto que no te aporten otros.

# Objetivos

El objetivo general de este proyecto consiste en desarrollar un sistema de Inteligencia Artificial (IA) con un enfoque multimodal basado en arquitecturas de *Deep Learning,* con el fin de detectar automáticamente *DeepFake Voice* y analizar el contenido de los mensajes para identificar posibles estafas telefónicas, contribuyendo así a la prevención dedelitos informáticos.

## Objetivos específicos

Para alcanzar el objetivo general mencionado, se plantean los siguientes objetivos específicos:

1. Desarrollar un sistema para la detección de audio en tiempo real que permita la recolección de señales de voz para su posterior procesamiento.
2. Diseñar e implementar tanto el modelo de detección de voces sintéticas como el sistema de análisis de contenido fraudulento.
3. Integrar ambos módulos en una arquitectura multimodal que combine tanto audio como texto.
4. Entrenar y validar el modelo utilizando métricas como accuracy o F1-score.
5. Analizar las implicaciones éticas e impacto social derivados del uso de IA en comunicaciones de voz.

# Desarrollo de la solución técnica

Aquí va todo lo duro técnico

Descripción de datos, preprocesamiento, arquitecturas, entrenamientos, etc…

# Resultados

Métricas, comparación de resultados, casos de éxito y fallos, visualizaciones, etc…

# Implicaciones éticas e impacto social

El siguiente análisis comprende el marco jurídico legal que acompaña este proyecto, pues al estar escuchando una llamada personal, se deben conocer las bases que legitiman dicho tratamiento de datos.

Bajo el prisma de la normativa española y europea, debemos ceñirnos a los siguientes puntos:

1. Secreto de las comunicaciones (Constitución Española, Art. 18.3): Garantiza el secreto de las comunicaciones, especialmente las postales, telegráficas y telefónicas, salvo mediante resolución judicial. Esto cambia si eres un participante en la conversación, no es delito que uno de los interlocutores grabe la conversación.

2. Protección de datos personales (RGPD y LOPDGDD): La voz de una persona es un dato personal, ya que se puede identificar con ella. Aunque las grabaciones no se almacenen en ningún lado, sigue siendo un tratamiento de datos personales. Para poder trabajar con esto, amparados por una base legal, nos podríamos ceñir a lo siguiente:

a. Consentimiento: La persona que llama debe ser informada de que su voz va a ser analizada y con qué propósito. Según el Artículo 6.1ª del RGDP, debe ser libre, específico, informado e inequívoco, materializándose mediante una locución informativa clara y concisa al inicio de la llamada. Debe incluir como mínimo la identidad del responsable del tratamiento y la finalidad explícita, así como información sobre cómo ejercer sus derechos (acceso, rectificación…), teniendo así una política de privacidad.

b. Interés legítimo: Frente a un juez, se tendría que justificar “El Juicio de Ponderación”, donde se pesa el interés contra los derechos de las personas que nos llaman, teniendo así que justificar:

i. ¿Es necesario?

ii. ¿Es proporcionado?

iii. ¿Prevalece de interés?

Es decir, se actuaría sin permiso directo. Existe el riesgo de que la base jurídica del tratamiento sea cuestionada, y es que, si una autoridad considera en el balance de intereses que el derecho de protección de datos del usuario prevalece sobre el interés de seguridad, no se justificaría por este método, y para cumplir con las normativas vigentes, se debería actuar por el método 2.a.

Si en un futuro este modelo se pone en producción, existen otros reglamentos importantes que se deben tener en cuenta, sobre todo con la llegada de la nueva normativa de la Inteligencia Artificial:

3. Reglamento de Inteligencia Artificial de la Unión Europea: Ha sido aprobado en 2024, contemplando los diferentes tipos de sistemas de Inteligencia Artificial según su nivel de riesgo. Esta ley obliga a las personas afectadas por dichos sistemas a ser informadas de que están interactuando con un sistema de Inteligencia Artificial

4. Directiva sobre la privacidad y las comunicaciones electrónicas: Otra ley específica complementando al RGDP, enfocada en el sector de las comunicaciones electrónicas y su confidencialidad, prohibiendo escuchas, rastreos o almacenamiento por parte de terceros sin el consentimiento.

5. Normativas de Ciberseguridad (Directiva NIS2 y Esquema Nacional de Seguridad): Obliga a tener diligencia debida en la seguridad de la aplicación, pues al final se están tratando datos biométricos de la persona que llame. En caso de ciberataque, se debe asegurar que nadie pueda acceder sin permiso a los resultados, o manipular dicho modelo de Inteligencia Artificial para que dé resultados erróneos.

Estos últimos reglamentos mencionados, refuerzan directamente el consentimiento, y es que, la Ley de IA convertirá en una obligación legal explícita notificar al interlocutor. Esto supone que argumentar “interés legítimo” sin informar, será una práctica inviable. Por lo que, para cumplir el marco legal, el interlocutor debe ser notificado al inicio de cada llamada escuchando lo siguiente:

“Le informamos que esta llamada puede ser grabada con el fin de analizar la voz por motivos de seguridad y prevención de fraude con un modelo de Inteligencia Artificial. Si continua en la llamada acepta dicho análisis”.

Bajo el marco de este tipo de aplicaciones, un factor para tener en cuenta son los riesgos de falsos positivos y negativos. De la mano de la normativa de Ciberseguridad y los riesgos operacionales, el alertar constantemente sobre llamadas legítimas, generaría desconfianza sobre la aplicación, lo que supondría su desuso debido a una fatiga de alertas. Sin embargo, el peor de los casos se encuentra con los falsos negativos, es decir, clasificar como voz legítima a una voz generada por IA. Esto supondría una falsa sensación de seguridad, volviendo más vulnerable al usuario, facilitando por tanto el ataque. Si dicho ataque se materializa, la confianza en la fiabilidad se destruye por completo, aun advirtiendo de las limitaciones de la IA, además de las pérdidas económicas y robo de datos personales asociado. Por ello, se debe asegurar la integridad, disponibilidad y resiliencia de los sistemas como indican las normativas, aplicando una Gestión de Riesgos continua (Identificar, Analizar, Implementar medidas y Monitorizar).

El uso adecuado de esta herramienta no es más que un impacto positivo en la sociedad. Sobre todo, en la protección de colectivos vulnerables, pues no suelen poseer conocimientos tecnológicos avanzados lo que supone el desconocimiento de las nuevas estafas y sus capacidades. Este tipo de estafas es una industria criminal masiva, pues los estafadores pueden llegar a suplantar tanto a un familiar como a bancos y compañías eléctricas robando credenciales con el fin de conseguir compensaciones económicas.

# Conclusiones

Que se ha logrado y que no. [Tabla de objetivos específicos si están cumplidos o no? ]

Futuras mejoras

# Bibliografía

Aquí se añaden las citas en APA.

# Anexos