SEMANA DE LA ADMINISTRACIÓN ABIERTA

DEEPFAKES Y FRAUDES BASADOS EN INTELIGENCIA ARTIFICIAL

MAYO 2026

BANCO DE ESPAÑA
Eurosistema

1

Introducción a la IA

Inteligencia artificial generativa.

IA GENERATIVA: UNA BREVE INTRODUCCIÓN

ChatGPT

Copilot

Gemini

Eleven Labs

Siri

Grok

LA IMPORTANCIA DE ELEGIR BIEN

Aplicaciones positivas

- Protección contra fraudes.
- Detección de patrones sospechosos.
- Análisis de comportamiento para la detección del fraude.

Desafíos éticos

- Privacidad.
- Acceso a grandes cantidades de datos.
- Gran sofisticación.
- ¿Cuándo confiar?

Aplicaciones negativas

- Phishing avanzado y automatizado.
- Deepfakes.
- Suplantaciones de identidad (por vídeo, por voz, etc.).

2

Fraudes basados en IA

Un cambio en la calidad de los engaños digitales.

¿QUÉ TIPO DE FRAUDES NOS ENCONTRAMOS?

Según el canal utilizado

Phishing
Correo Electrónico

Smishing
SMS y Mensajería Instantánea

Vishing
Llamada telefónica

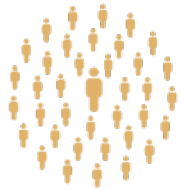
Angle Phishing
Redes Sociales

QRishing
QR

Pharming
Web

Evil Twin Phishing
Redes WIFI

Según el objetivo

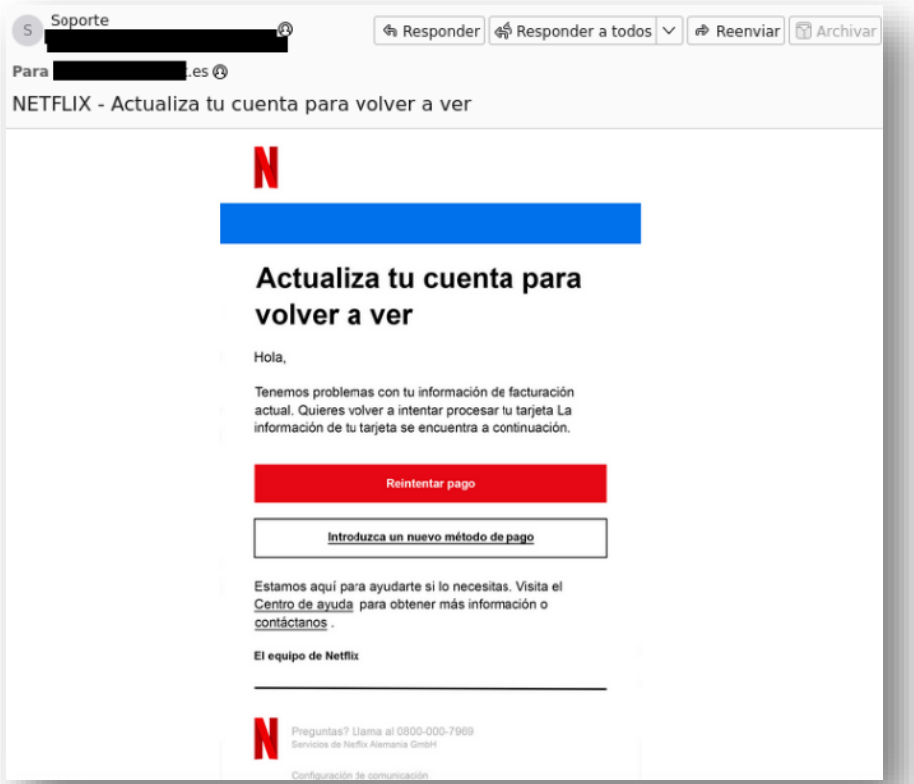


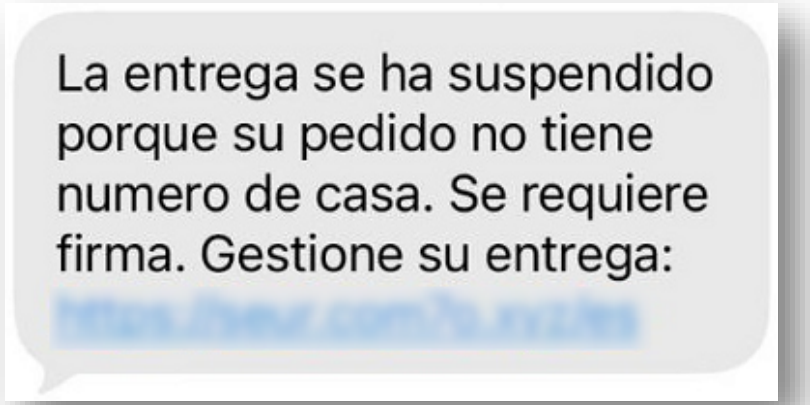
Phishing masivo
Impersonal

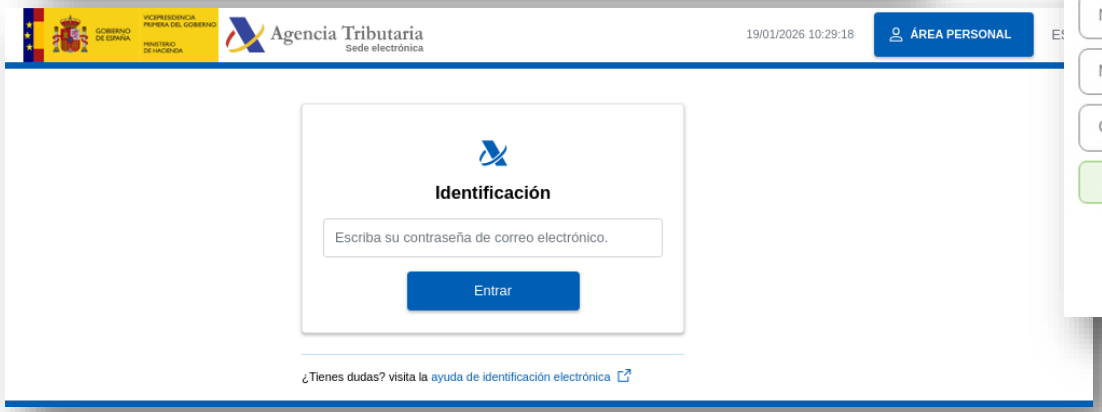
Spear phishing
Dirigido a una persona

Whaling
Dirigidos a ejecutivos de alto nivel

¿QUÉ TIPO DE FRAUDES NOS ENCONTRAMOS?







¿QUÉ SON LOS “DEEPFAKES”?

Vídeos, **imágenes** o **audios** creados o modificados con **Inteligencia Artificial**

USOS ÉTICOS:

- ☑ **Medicina y accesibilidad**
- ☑ **Entretenimiento y cine**
- ☑ **Educación y cultura**

¿QUÉ SON LOS “DEEPFAKES”?

USOS FRAUDULENTOS:

- ☒ **Suplantación** de identidades, **ciberacoso**
- ☒ **Manipulación** de opiniones y desinformación
- ☒ **Estafas** de ingeniería social

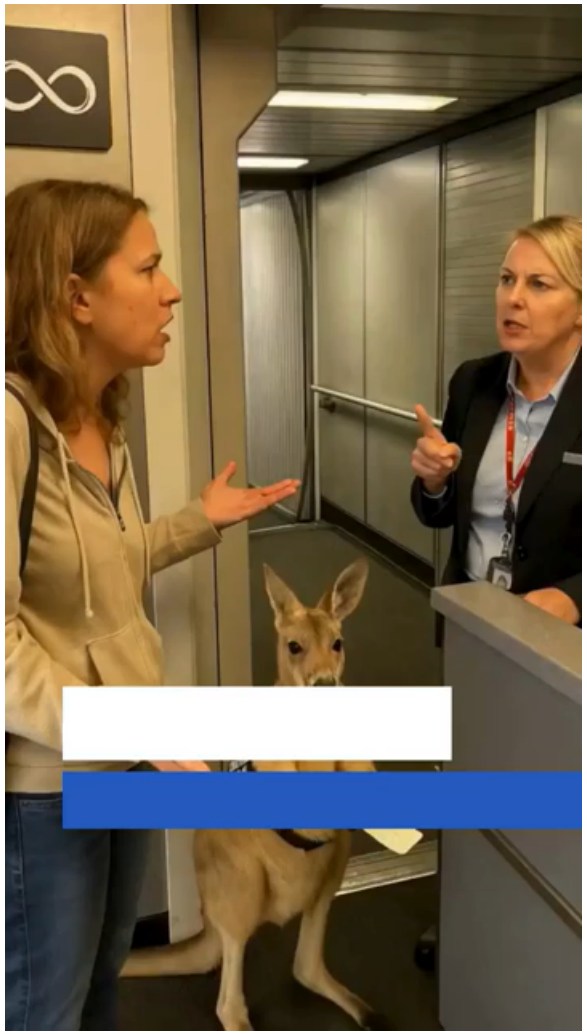
8M	+3000%	$500k	24,5%
deepfakes en 2025	fraudes en 2023	pérdida media por incidente	tasa de detección humana

Fuente: https://deepstrike.io/blog/deepfake-statistics-2025

3

Esto ya no es ciencia ficción

Un acercamiento a las técnicas utilizadas.



VerificaRTVE

PROBLEMA DE CIBERSEGURIDAD

Estafa con 'deepfakes': 24 millones de euros y una reunión virtual con solo una persona real

• Una videoconferencia entre un empleado y varios ejecutivos de una multinacional británica resultó ser una estafa que ha supuesto pérdidas millonarias para la empresa

CIBERSEGURIDAD >

Los usos criminales de los 'deepfakes' se disparan: estafas, pornografía y suplantación de identidad

Los expertos y las compañías de ciberseguridad advierten de que esta tecnología ahora es capaz de engañar a algunos métodos de autenticación biométrica o crear perfiles de trabajadores falsos que postulan a puestos reales

Los padres de un menor sancionados por la difusión de imágenes falsas de chicas desnudas generadas con IA en Almendralejo

La multa, de 2.000 euros, se impone por la implicación de su hijo en la creación y difusión de imágenes manipuladas con inteligencia artificial que afectaron a varias adolescentes de la localidad pacense

Crecen denuncias de suplantación y no tienes que ser CEO o influencer para ser víctima

Se eleva uso de inteligencia artificial, y ahora hay históricas sentencias por este delito en TikTok y Facebook

DELITOS CONTRA LA INTIMIDAD

Recuerda:

1. No uses la imagen de otras personas con la IA sin su consentimiento.
2. No hagas nada que no te gustaría que te pasase.
3. Piensa en las consecuencias: repercusiones legales y daños psicológicos.

El Gobierno aprueba una nueva Ley del derecho al honor más garantista y adaptada al entorno digital

13.1.2026

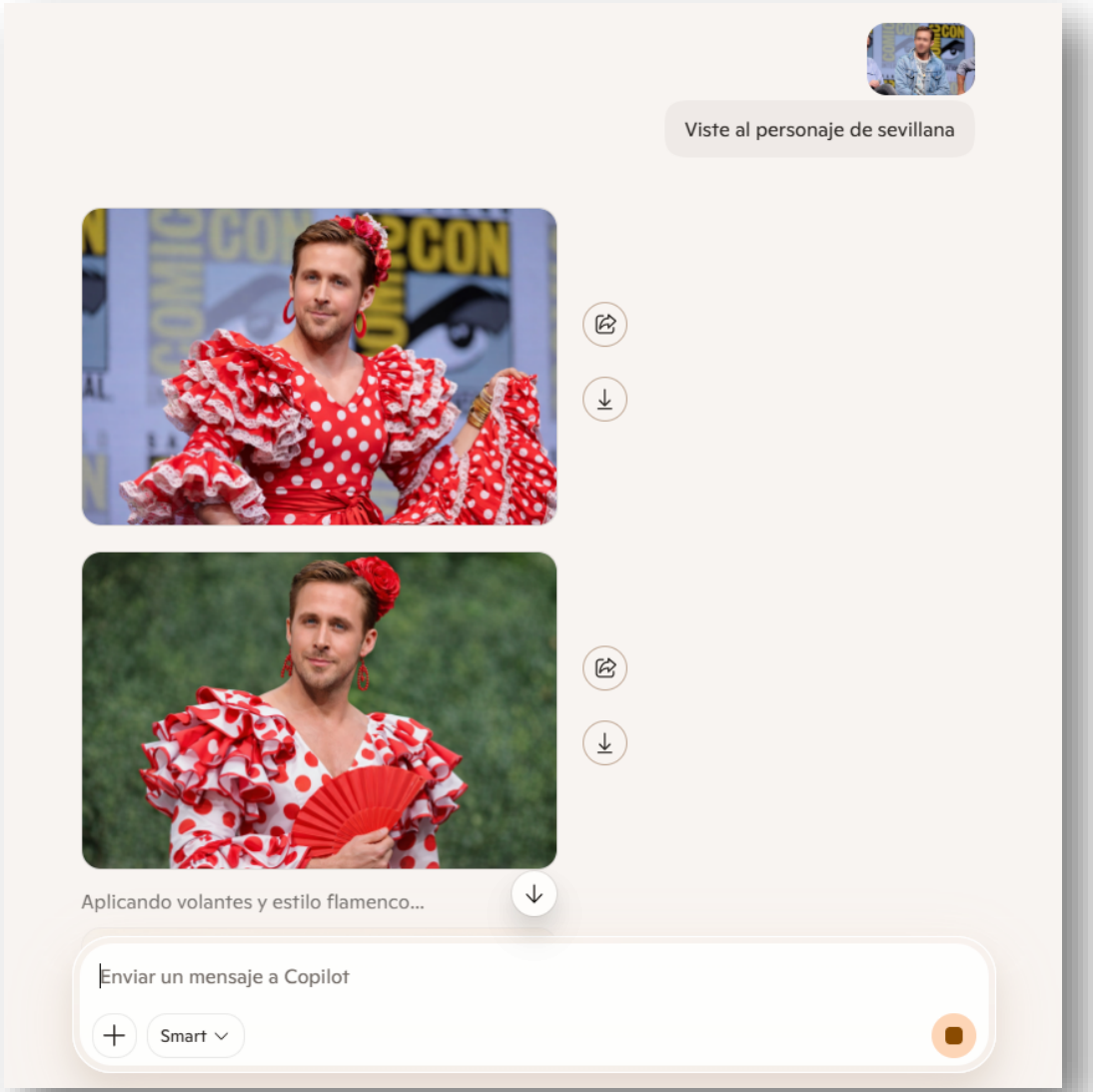
El Gobierno aprueba tipificar como delito los deepfakes de contenido sexual y el grooming

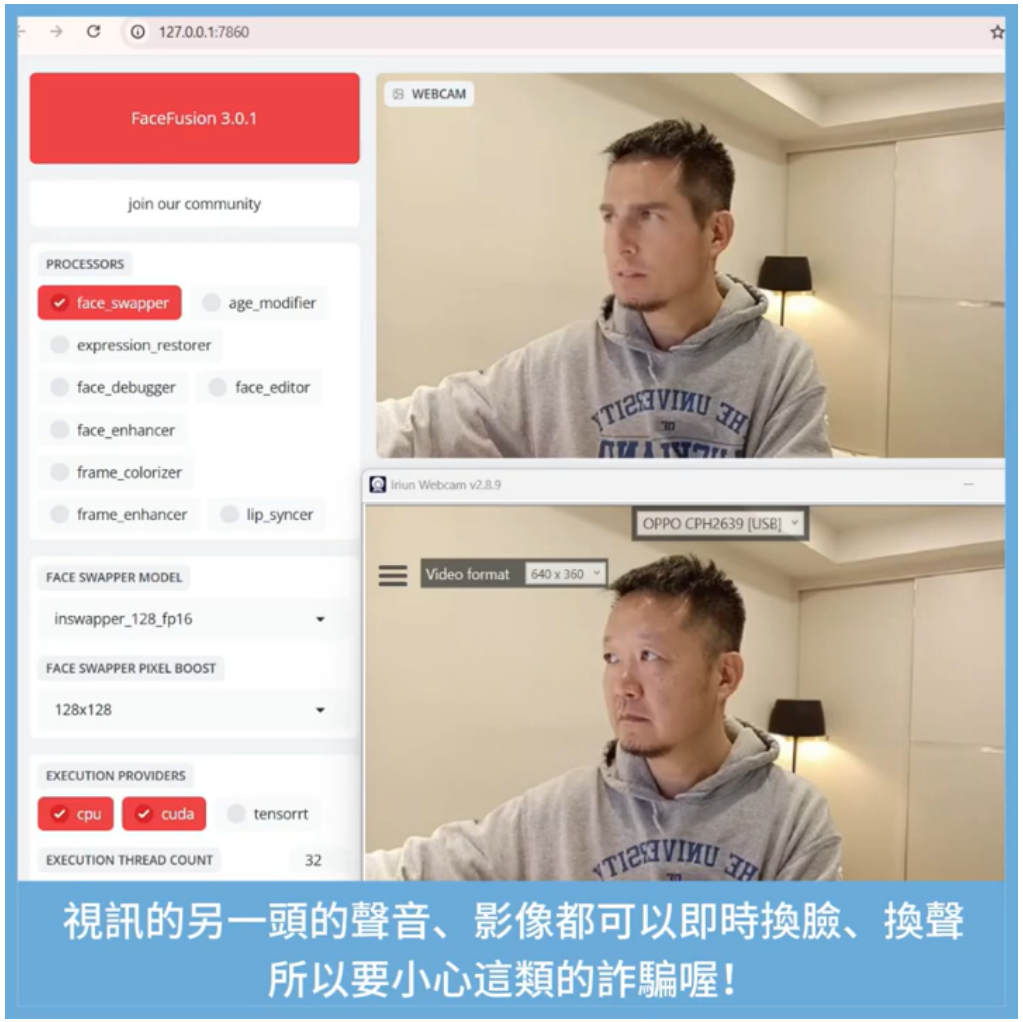
La Moncloa Castellano

La Unión Europea acuerda la propuesta de España de prohibir los 'deepfakes' sexuales

Transformación Digital y Función Pública - 7.5.2026

Crear o difundir imágenes íntimas generadas por IA sin consentimiento no es una broma tecnológica. **Es un delito**.





4

DEEPVOICE

Caso práctico y sus repercusiones.

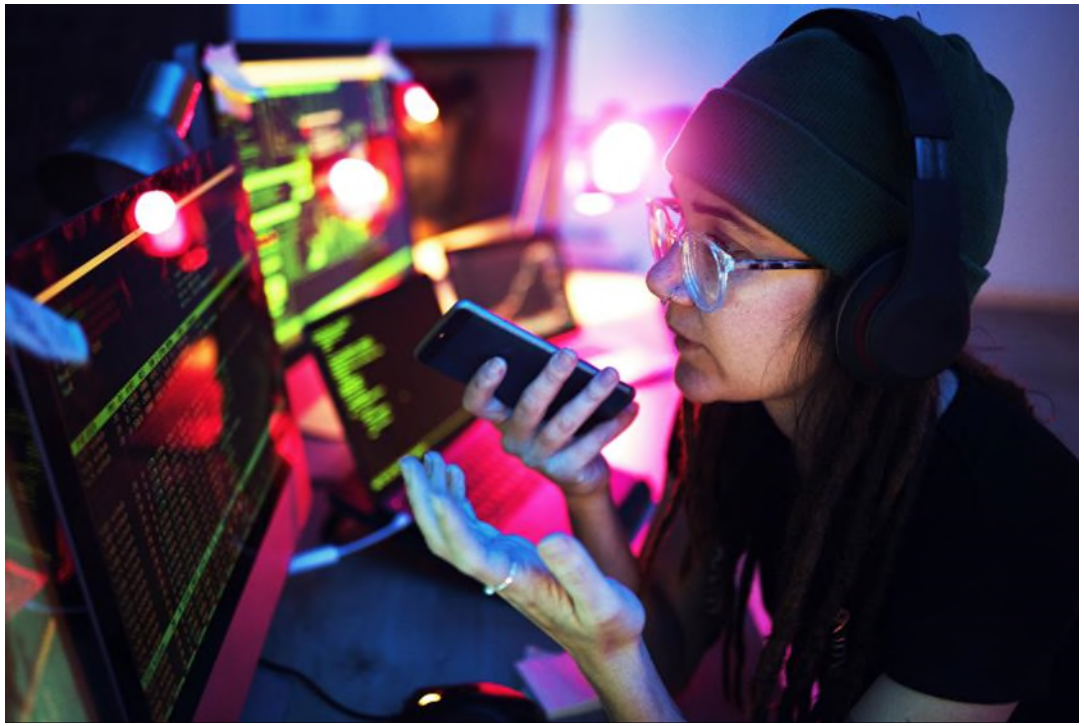
BENEFICIOS DE LA CLONACIÓN DE VOZ

- **Accesibilidad**
- **Industria del entretenimiento (cine, videojuegos…)**
- **Asistentes virtuales**
- **Audiolibros**
- **Doblaje y traducción**
- **Detección del fraude**

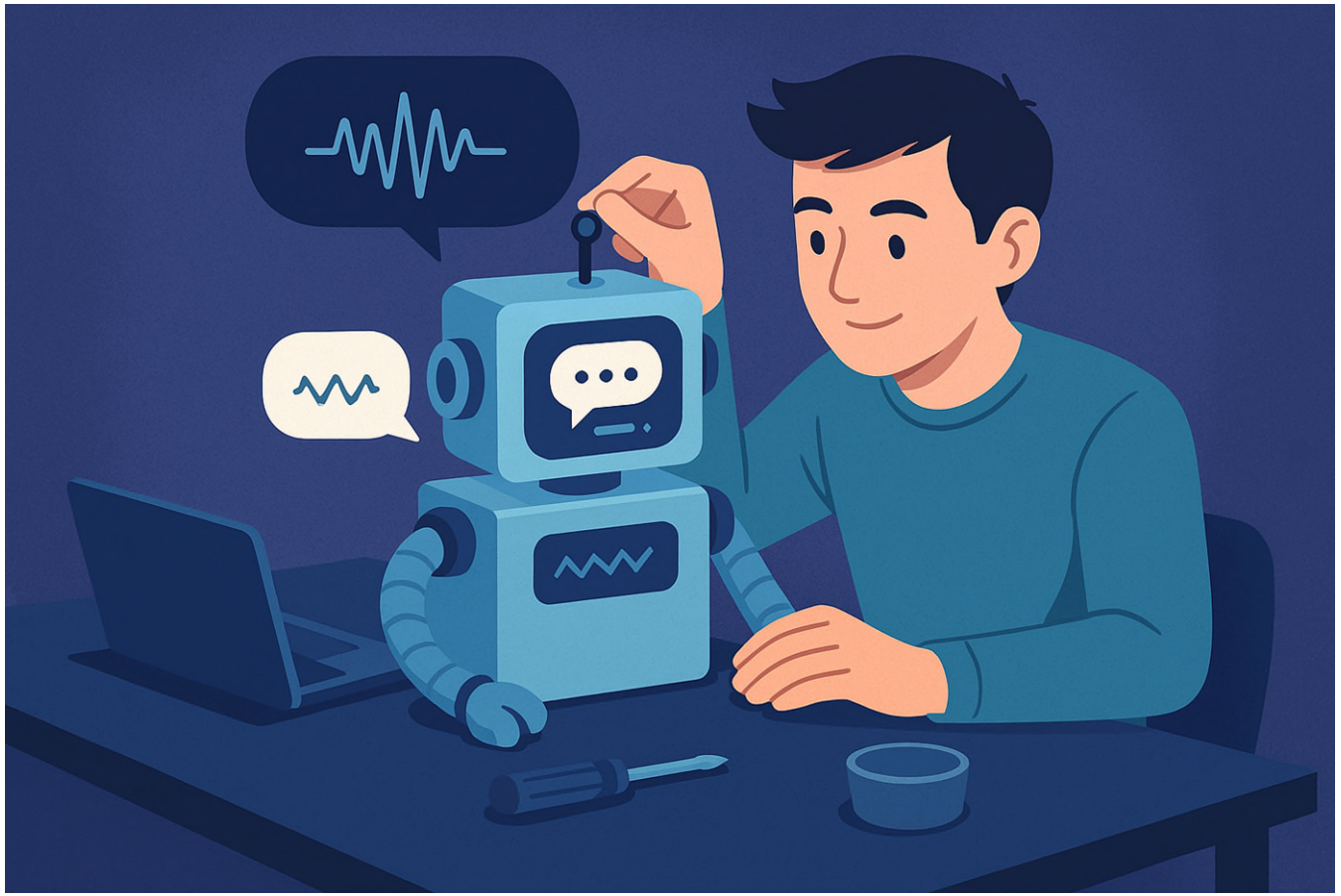
Una **Inteligencia Artificial** que **permite hablar** a quien no tiene voz, que dinamiza la industria del **entretenimiento** y que sirve para detectar **fraudes**.

USO FRAUDULENTO DE LA CLONACIÓN DE VOZ

- **Fraudes y estafas**
- **Manipulación de la información**
- **Vulneración de la seguridad**
- **Pruebas falsas**
- **Falsos consentimientos**
- **Consecuencias psicológicas**



DIY: HAZLO TÚ MISMO

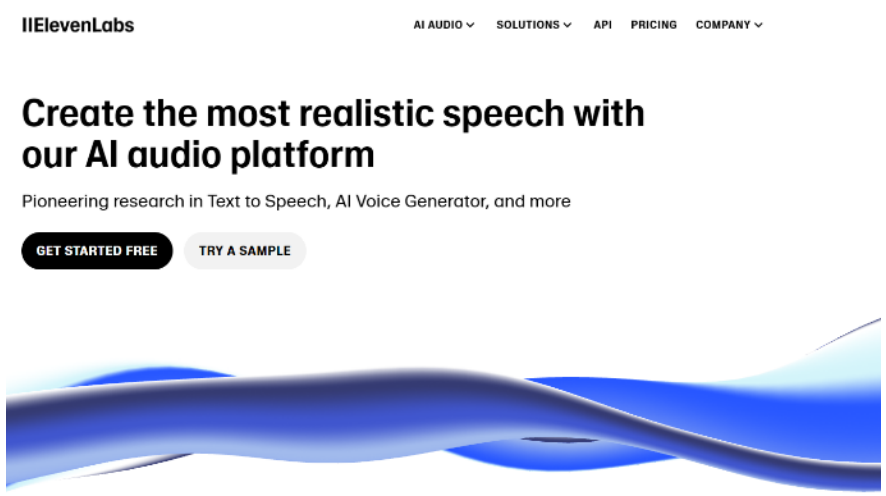


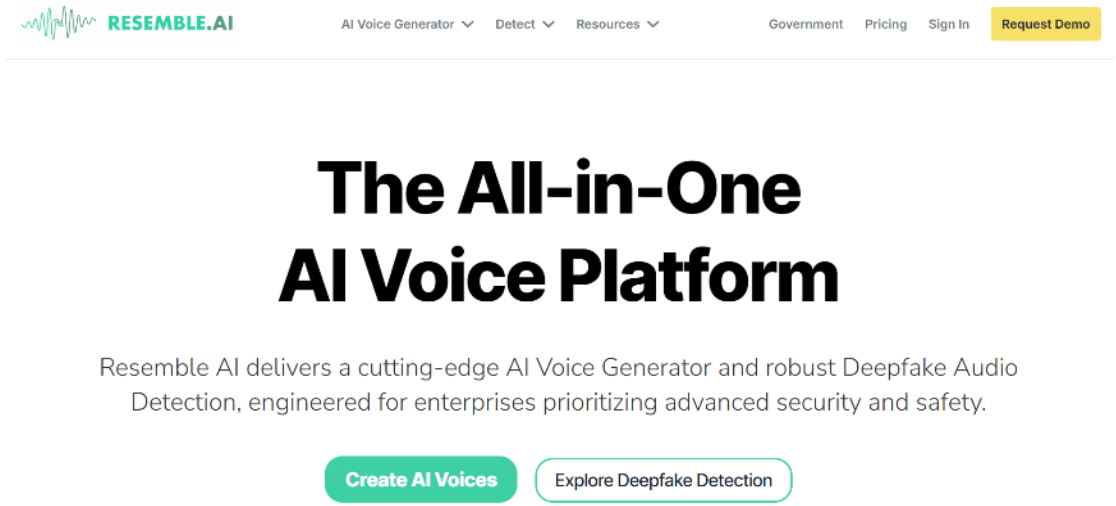
LAS TRES FASES DE LA CLONACIÓN DE VOZ

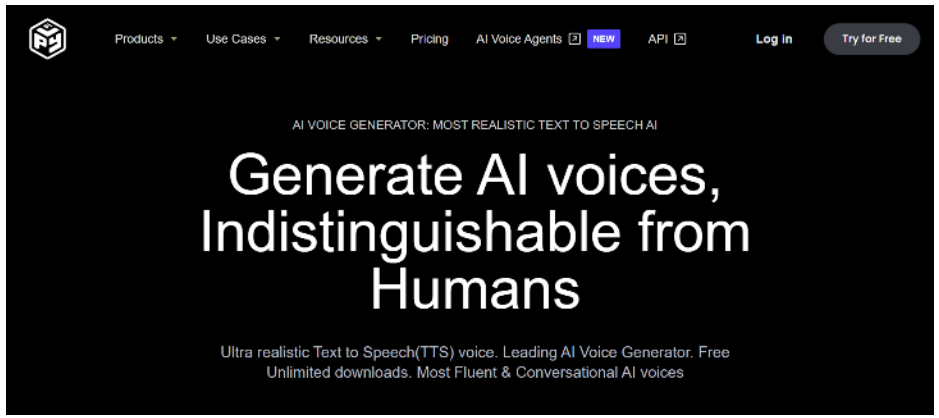
1. **Recolección de datos**: grabación de la voz original.
2. **Entrenamiento del modelo**: pronunciación, estilo.
3. **Creación de voz**: nuevas frases.

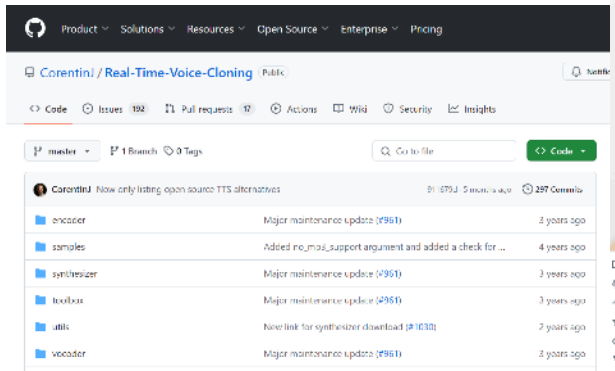
El deep learning utiliza redes neuronales para aprender y reconocer patrones en grandes conjuntos de datos.

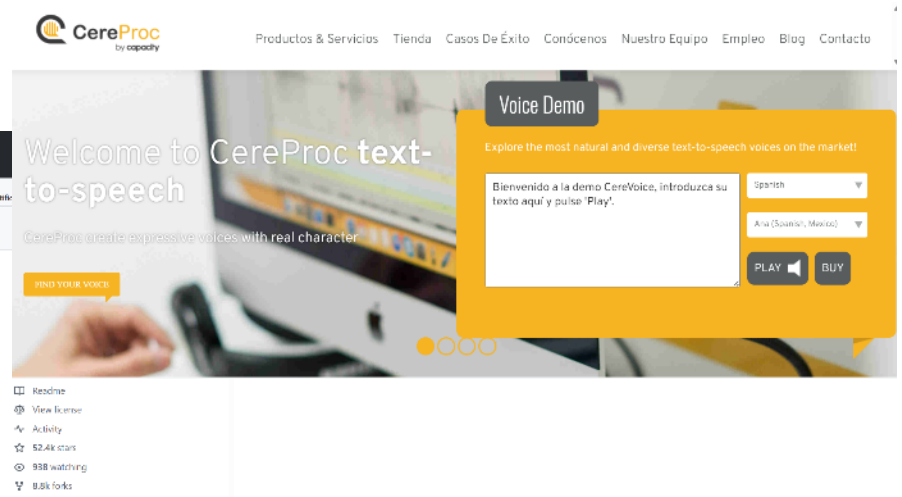
MÉTODOS DISPONIBLES









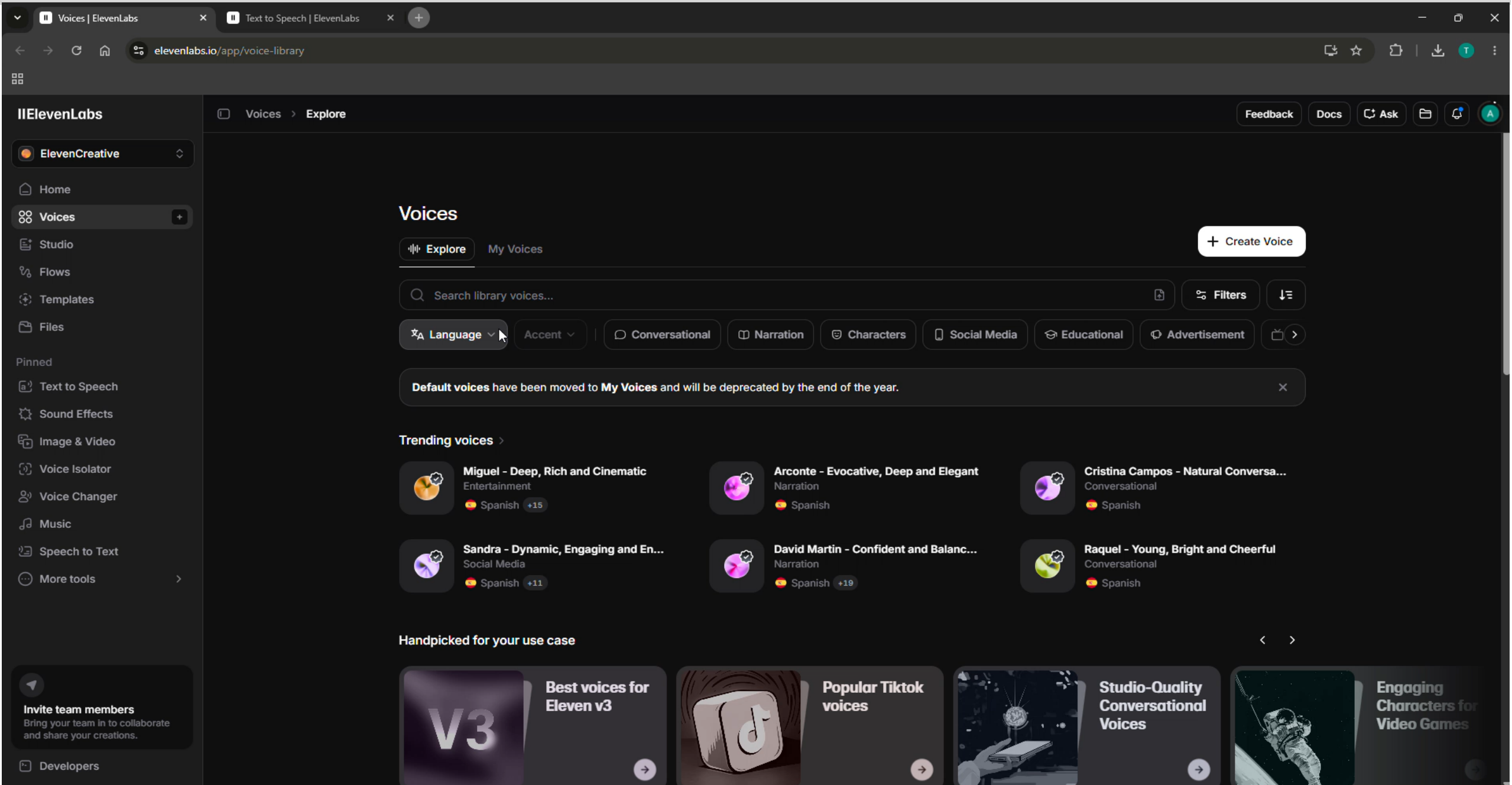


Departamento IT 09:16

Buenos días. Su cuenta aparece en el sistema como el origen de un ciberincidente. Su conexión puede verse afectada, le mandamos las instrucciones por audio para seguirlas paso a paso.

Hola, soy el responsable de la unidad. Necesito que envíes una documentación lo antes posible, te mando la información en un audio.

Chat Screenshot by pranx.com



5 CONCLUSIONES

IA: Una frontera ética.

USA LA IA DE FORMA SEGURA

Cuida tu privacidad y evita el uso de información personal:

- Limita quién puede ver tus vídeos, fotos o audios en redes sociales.
- Evita facilitar esta información en aplicaciones de IA generativa.

Analiza los contenidos que consumas:

- Verifica su veracidad a través de aplicaciones y fuentes externas.
- Busca errores en los contenidos, como parpadeos, manos y gestos extraños.
- Verifica la identidad de las personas a través de otras vías si te interpelan personalmente.

Desconfía de contenidos extraños o alarmantes:

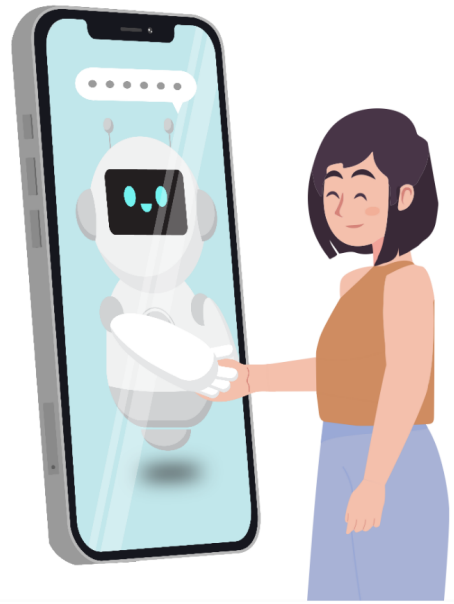
- Vídeos, imágenes y audios sensacionalistas.
- Mensajes demasiado buenos para ser verdad o que apelen a las emociones.

Mantente alerta ante la ingeniería social:

- Desconfía por defecto de peticiones inusuales o de desconocidos.
- Reporta cualquier anomalía o sospecha de suplantación de identidad.

Reporta el contenido fraudulento generado con IA:

- Cuando usen tu imagen o la de una persona conocida de tu entorno.
- Usa los métodos de reporte a través de los mecanismos de las redes sociales.
- Busca ayuda en organismos como INCIBE o la AEPD.
- Denuncia los delitos a las Fuerzas y Cuerpos de Seguridad del Estado.



¿CÓMO PODEMOS RECONOCER UN FRAUDE CON IA?

¿Suena igual que siempre?	**¿Cuántos dedos tienen las manos?**
¿Hay demasiadas pausas?	**¿Tiene expresiones faciales extrañas?**
¿Notas cambios de tono?	**¿Es correcta la sincronización?**
¿Hay fallos en la imagen?	**¿Aparecen fallos en el fondo?**

RECUERDA: también minimizamos riesgos evitando mensajes inusuales y **comprobando siempre la información**.

PASOS A SEGUIR PARA DETECTAR VÍDEOS CON IA

1. Localiza el origen de la publicación.
2. Amplía la información con fuentes externas.
3. Analiza el vídeo y el audio con detenimiento.
4. Comprueba los metadatos.
5. Utiliza los detectores de contenidos generados con IA.

Herramientas útiles para identificar contenido falso generado por IA

Para complementar los consejos facilitados anteriormente, es esencial emplear técnicas prácticas que permitan una verificación más profunda y técnica del contenido. A continuación, se describen algunas herramientas que pueden servir de ayuda:

1. **Detectores de IA para texto**: Existen varias herramientas diseñadas para analizar textos y detectar si han sido generados artificialmente. Mostramos algunos ejemplos:
   1. plagiarismdetector: analiza patrones lingüísticos y estructuras gramaticales para indicar un tanto por ciento de posibilidades de que sea generado artificialmente.
   2. GPTZero: tiene la capacidad de detectar contenido generado por modelos como ChatGPT y GPT-4. Muestra los porcentajes de probabilidad de autoría humana o IA, y resalta las secciones sospechosas.
   3. Copyleaks: proporciona un análisis detallado a nivel de oración para detectar contenido generado por IA con una alta precisión.
2. **Herramientas de detección de multimedia**: Para detectar *deepfakes* y otros tipos de contenido multimedia generado por IA, estas herramientas pueden ser muy útiles:
   1. VerifAI, una web respaldada por la empresa Telefónica que detecta en pocos segundos si un contenido ha sido generado o alterado con IA.
   2. Resemble Detect detecta audios generados con IA.
   3. Deepware: permite analizar vídeos y detectar deepfakes mediante un análisis exhaustivo que identifica alteraciones y anomalías.
   4. Illuminarty: proporciona análisis avanzado para detectar deepfakes en imágenes, utilizando inteligencia artificial para identificar discrepancias y manipulaciones.
   5. AI or Not: permite a los usuarios cargar imágenes y audios para verificar si han sido manipulados o generados por inteligencia artificial.
   6. VerificAudio: ayuda a verificar, enviando a un equipo especializado, voces sintéticas.

incibe_ INSTITUTO NACIONAL DE CIBERSEGURIDAD

OSI Oficina de Seguridad del Internauta

La mejor herramienta es el sentido común

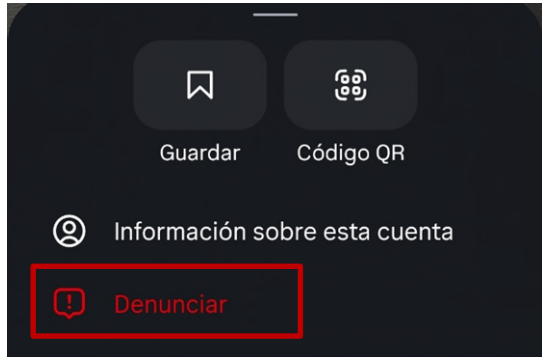
REPORTA EL CONTENIDO EN REDES SOCIALES

facebook

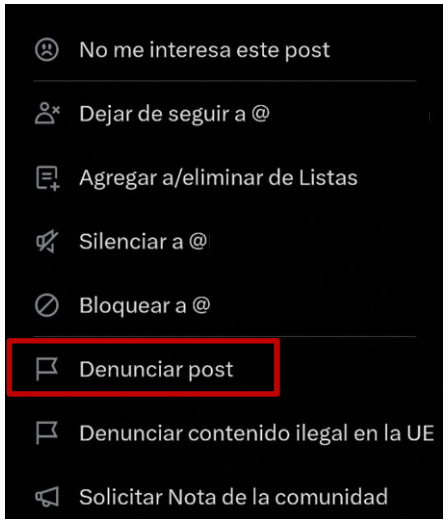
(En los tres puntos)

Instagram

(En los tres puntos)

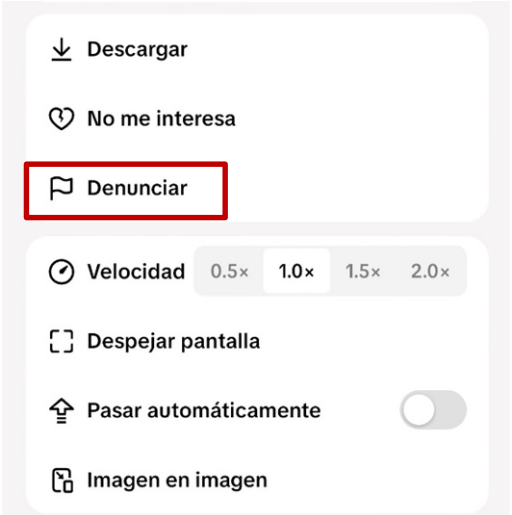


(En los tres puntos)



TikTok

(Dejando pulsado)



CONTACTA CON EL INCIBE (INSTITUTO NACIONAL DE CIBERSEGURIDAD)

CONTACTA CON LA AEPD (AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS)

Formularios de solicitudes para el ejercicio de los derechos

- Formulario de solicitud para el ejercicio del derecho de acceso
- Formulario de solicitud para el ejercicio del derecho de rectificación
- Formulario de solicitud para el ejercicio del derecho de oposición
- Formulario de solicitud para el ejercicio del derecho de supresión
- Formulario de solicitud para el ejercicio del derecho a la limitación del tratamiento
- Formulario de solicitud para el ejercicio del derecho a la portabilidad de los datos

DENUNCIA EL CASO A LAS FUERZAS Y CUERPOS DE SEGURIDAD DEL ESTADO

Policía Nacional	Guardia Civil	Ertzaintza	Mossos d'Esquadra	Policía Foral
Brigada de Investigación Tecnológica de la Unidad de Investigación Tecnológica (UIT) de la Policía	Grupo de Delitos Telemáticos (GDT) de la Guardia Civil	Sección Central de Delitos en Tecnologías de la Información (SCDTI) de la Ertzaintza	Unidad Central de Delitos Informáticos de los Mossos d'Esquadra	Grupo de Apoyo Tecnológico de la Policía Foral de Navarra

Muchas gracias por su atención

BANCO DE ESPAÑA
Eurosistema