Como saber si un video es falso hecho con IA

@StephanoBrooke · Palabra clave: FALSO

Lo que necesitas

Nada especial: tu celular, ChatGPT (sirve gratis) y Google Lens, que ya viene en la app de Google y en Google Fotos.

1. Revisa quien lo publico

Antes de creerlo o compartirlo, revisa la cuenta:

- ¿Es la cuenta oficial? Mira la verificación, la antigüedad y si sus publicaciones anteriores tienen sentido.
- ¿Es sensacionalista? Si te pide urgencia, miedo o dinero, desconfía.
- ¿Algún medio serio lo publicó? Si es noticia de verdad, varios medios ya la tienen.

2. Busqueda inversa

- Haz captura de la imagen o de un cuadro del video.
- Ábrela con Google Lens (en la app de Google o en Google Fotos, botón de Lens) y elige buscar la imagen.
- Si aparece en otros sitios con otra fecha o con otro contexto, es reciclada o está editada.

También puedes subirla a TinEye.com desde la computadora.

Prompt 1: Revisar un mensaje sospechoso

> *"Me llegó este mensaje o video: [pega el texto o descríbelo]. Lo mandó [quién y por dónde]. Dime qué señales de estafa o de contenido falso ves, qué debo verificar antes de creerlo y dónde puedo confirmarlo. Dímelo en una lista corta."*

Que pasa despues: una lista con las señales de alerta y los pasos para confirmar antes de responder o compartir.

Prompt 2: Llamada o audio de un familiar

"Me escribió o me llamó alguien que dice ser [mi hijo, mi mamá, un amigo] y me pide [dinero, un depósito, un código]. Dice que es urgente porque [lo que te dijo]. Dime qué preguntas hacerle para confirmar que es esa persona y qué no debo hacer por ningún motivo."

Que pasa despues: preguntas de verificación que solo esa persona sabría y lo que nunca debes hacer (depositar o pasar códigos sin confirmar).
Regla de oro: cuelga y marca tú al número que ya tienes guardado de esa persona.

3. La palabra clave familiar

Es una palabra que solo saben en tu familia. Si alguien llama con la voz de tu hijo pidiendo dinero, le pides la palabra. Una voz clonada con IA no la sabe.

- Que sea fácil de recordar y difícil de adivinar.
- Que no aparezca en sus redes (ni nombres de mascotas ni fechas).
- Acuérdenla en persona o por llamada, no por un chat que se pueda hackear.
- Cámbienla si alguien la usó o si se la dijeron a alguien de fuera.

Prompt 3: Ideas de palabra clave

"Dame 10 ideas de palabras clave familiares que sean fáciles de recordar para personas mayores, difíciles de adivinar y que no tengan relación con nombres, fechas o mascotas."

Que pasa despues: una lista de opciones para escoger en familia.

Prompt 4: Mensaje para el grupo familiar

"Escríbeme un mensaje corto y cariñoso para el grupo de WhatsApp de mi familia explicando que ya existen llamadas y videos falsos con IA que imitan voces, y que vamos a usar una palabra clave para confirmar cualquier pedido de dinero. Que no asuste y que lo entienda alguien de 70 años."

Que pasa despues: un mensaje listo para copiar al grupo de la familia.
Ojo: no pongas la palabra clave en el mensaje. Esa se acuerda en persona o por llamada.

Senales de un video falso

- Boca y dientes que no coinciden con lo que dice, o dientes borrosos.
- Manos, dedos, lentes o aretes que se deforman al moverse.
- Parpadeo raro y mirada fija.
- Bordes de la cara que se ven pegados o que tiemblan.
- Texto o letreros del fondo que no se leen bien.
- Voz sin respiraciones, sin ruido de fondo o con tono plano.

Ninguna señal sola lo confirma, y los videos son cada vez mejores. Por eso la verificación por otra vía siempre gana.

Prompt 5: Analizar una imagen

> *"Analiza esta imagen [súbela]. Dime qué partes podrían indicar que está hecha o editada con IA (manos, texto, sombras, reflejos, bordes) y qué tan seguro estás. Si no puedes saberlo, dímelo."*

Que pasa despues: te señala detalles sospechosos de la imagen y su nivel de certeza.
ChatGPT puede equivocarse en las dos direcciones. Úsalo como una opinión más, no como prueba.

Prompt 6: Verificar una noticia

> *"Busca si esta noticia es real: [pega el titular o el texto]. Dime qué medios confiables la publicaron, desde cuándo, y si hay desmentidos. Pon los links de tus fuentes."*

Que pasa despues: si tiene búsqueda, te dice qué medios la confirman o la desmienten, con links para revisar tú.

Si ya caiste: Mexico

- Tu banco primero: llama o entra a la app, bloquea tarjetas o cuentas y pide folio de la aclaración.
- Protege tus cuentas: cambia contraseñas, cierra sesiones y activa la verificación en dos pasos.
- Guarda evidencia: capturas con fecha, números, links, comprobantes. No borres nada.
- Corta el contacto y no compartas más datos.
- CONDUSEF: si el banco no te resuelve, en su Portal de Fraudes Financieros.
- Policía Cibernética de tu estado o el 088 de la Guardia Nacional para delitos en línea.
- 089 para denuncia anónima de extorsión, y 911 si hay riesgo inmediato.
- Ministerio Público si quieres levantar denuncia formal.

Prompt 7: Plan después de una estafa

> *"Me acaban de estafar así: [qué pasó, sin poner contraseñas ni números de tarjeta]. Estoy en México. Dame los pasos en orden para hoy: a quién llamo primero, qué evidencia guardo y qué cuentas protejo."*

Que pasa despues: una lista en orden para actuar rápido sin olvidar nada.
Nunca le pegues a ChatGPT tus contraseñas, NIP ni números completos de tarjeta.

Prompt 8: Crear cada personaje en Higgsfield

El reel se graba de corrido en un solo lugar y en Higgsfield se cambia el personaje frase por frase con control de movimiento. Todos los personajes son inventados. Prompt (reconstruido viendo el video original):

> *"Fotografía realista, plano medio, una persona de pie en el centro de un lobby amplio de aeropuerto con piso brillante y ventanales al fondo, luz de día, mirando a cámara y hablando con las manos. Personaje: [PERSONAJE]. Persona inventada, no se parece a ninguna celebridad."*

Que pasa despues: una imagen por personaje para usar en Higgsfield.

Como usarlos

Hoy: acuerda la palabra clave (prompts 3 y 4). Cada vez que algo te huela raro: prompts 1, 2, 5 y 6. Si ya caíste: la lista de México y el prompt 7.
Mándale esta guía a quien más la necesite y cuéntame en el reel si tu familia ya tiene su palabra clave. Etiquétame en @stephanobrooke.