IA para el día a día: Seguridad

7 guías paso a paso para tareas reales del día a día

Seguridad

Actualizado: 8 de noviembre de 2026

AI World HQ · https://aiworldhq.com/es

Edición gratuita · CC BY-NC 4.0 · compártela citando la fuente

Antes de empezar: qué es esto, con sinceridad

Esta es una edición actualizada. Se rehízo a partir de nuestras guías actuales en la fecha que aparece en la portada: más guías por tema, sin las guías retiradas y con cada paso tomado de la versión más reciente de cada guía.

Esta colección es gratuita, y todas sus guías están también gratis en nuestra web. Este archivo solo las reúne en un lugar, en orden de lectura, para que puedas imprimirlo, llevarlo en el móvil o leerlo sin conexión. Nuestra web tampoco tiene muro de pago: cada guía enlaza a su página en internet, que siempre es la versión más reciente.

AI **Todas las guías de este paquete fueron escritas por IA** y pasaron controles automáticos de datos y calidad. **Nuestro editor comprueba cada día una muestra de lo publicado, pero no todos los textos los lee una persona.** Cada guía de https://aiworldhq.com lleva el mismo aviso: lo decimos abiertamente porque tienes derecho a saber quién escribió lo que estás leyendo.

Las aplicaciones cambian. Los botones se mueven y los menús cambian de nombre. Si un paso no coincide con lo que ves, la versión en internet de la guía es la más reciente: el enlace está debajo de cada título.

Puedes compartirla libremente. Esta colección tiene licencia Creative Commons Atribución-NoComercial 4.0 (CC BY-NC 4.0): puedes copiarla, imprimirla y compartirla —con tu clase, tu biblioteca o tu equipo— siempre que cites a AI World HQ con un enlace a aiworldhq.com y no la vendas.

Qué incluye

1. Seguridad

1. Cómo comprobar si un mensaje sospechoso es una estafa pegándolo en una IA
2. Cómo detectar hoy una llamada, un correo o una voz falsa generados con IA
3. Desactiva el historial y el guardado de datos en ChatGPT, Claude y Gemini
4. Lo que nunca deberías compartir con un asistente de IA: guía de privacidad
5. Cómo detectar un correo de phishing o un mensaje fraudulento antes de hacer clic
6. Cómo detectar estafas de IA, deepfakes y contenido falso
7. Cómo detectar un deepfake o una voz clonada antes de compartirlo

Seguridad

Cómo comprobar si un mensaje sospechoso es una estafa pegándolo en una IA

6 pasos · Léela en la web

> **En resumen:** Pega un mensaje sospechoso en cualquier asistente de IA — ChatGPT, Gemini, Claude, Copilot o Le Chat — y pídele que señale las señales de alerta. Obtendrás una segunda opinión en menos de un minuto, pero nunca pegues contraseñas reales y verifica siempre con la empresa primero.

Las estafas por mensaje se han vuelto muy buenas. Usan logotipos reales de empresas, tu nombre real y, a veces, incluso tu dirección real. Antes de hacer clic en nada, copia el mensaje en un asistente de IA y pide una segunda mirada. Esta guía te explica cómo hacerlo en seis pasos, incluso si nunca has usado un asistente de IA. Una advertencia honesta de entrada: la IA es una primera comprobación útil, no un veredicto, y no puede hacer clic en el enlace ni llamar a la empresa por ti. Esa parte sigue siendo tuya.

Antes de empezar

- Un asistente de IA con el que puedas iniciar sesión — ChatGPT, Gemini, Claude, Microsoft Copilot y Le Chat sirven, y sus planes gratuitos son suficientes para esto.
- El mensaje sospechoso listo para copiar (un correo, un SMS, una captura de pantalla de chat o un DM).
- Unos 5 minutos en total.
- Importante: nunca pegues contraseñas reales, números completos de tarjetas ni documentos de identidad, aunque el mensaje los contenga — en el paso 2 te explico cómo gestionarlo.

Paso 1 — Abre una conversación nueva en cualquier asistente de IA

Abre la web o la app de cualquier asistente de IA — ChatGPT, Gemini, Claude, Microsoft Copilot, Le Chat o el que ya uses. Empieza una conversación nueva en lugar de soltar el mensaje en un hilo antiguo. En el móvil, el icono de la app suele ser una burbuja de chat o un logotipo con estrellas. En el ordenador, ve a la página de inicio del asistente y haz clic en el cuadro de chat vacío. Un hilo limpio evita que la IA se distraiga con contexto anterior y se centre solo en tu mensaje. Si todavía no tienes cuenta, el registro gratuito suele tardar menos de un minuto y solo pide un correo electrónico. 💬 Ejemplo: abre ChatGPT, Gemini, Claude, Microsoft Copilot o Le Chat — el que ya tengas con cuenta. Sabrás que funcionó cuando veas un cuadro de chat vacío con un texto del tipo "Escribe un mensaje..." o "Pregunta lo que quieras" y un botón de enviar (a menudo una flecha o un icono de avión de papel).

Paso 2 — Tacha los datos personales antes de pegar

Antes de copiar el mensaje, tómate 30 segundos para ocultar cualquier dato sensible que contenga. Tapa contraseñas reales, números completos de tarjeta, tu número de identificación fiscal o de la seguridad social, tu domicilio y cualquier código de inicio de sesión de un solo uso. Un rápido buscar-y-reemplazar en cualquier app de notas funciona bien. Piensa en ello como tapar la mano al marcar el PIN en un datáfono: la IA no necesita esos datos para detectar una estafa, y cuantos menos números reales haya en el chat, menos riesgo habrá si tu cuenta se ve comprometida. Si el mensaje no tiene datos personales, puedes saltarte al paso 3. 💬 Ejemplo: convierte "mi DNI es 12345678X, llámame al 555-0123" en "mi DNI es [TACHADO], llámame al [TACHADO]". Sabrás que funcionó cuando el mensaje siga leyéndose de forma natural pero no quede ningún número, código ni dirección real.

Paso 3 — Pega el mensaje y pide a la IA que señale las señales de alerta

Toca o haz clic dentro del cuadro de chat vacío y pega el mensaje ya tachado. Luego añade una instrucción breve pidiendo al asistente que lo analice como si fuera un experto en seguridad. Pulsa el botón de enviar (normalmente una flecha ↗ o un avión de papel, aunque varía según la app). Tras unos segundos, la IA empezará a escribir una respuesta, por lo general una lista corta de señales de aviso más un veredicto de una línea como "esto parece un intento de phishing" (un mensaje falso diseñado para robarte el acceso o el dinero) o "parece legítimo, pero verifica al remitente". Si no aparece nada tras unos segundos, busca un pequeño icono de "detener" o un cuadrado: a veces la respuesta tarda un momento en empezar. Las apps cambian a menudo, así que si tu pantalla se ve algo distinta, busca el botón que envía el mensaje. 💬 Ejemplo: "Acabo de recibir este mensaje. ¿Puedes actuar como un experto en seguridad y decirme si parece una estafa? Enumera las señales de alerta concretas que ves y dime si debería hacer clic en algún enlace. Mensaje: [pega aquí]" Sabrás que funcionó cuando veas una respuesta escrita que mencione señales de aviso concretas, como urgencia, enlaces sospechosos, un remitente raro o peticiones de datos personales.

Paso 4 — Lee la respuesta de la IA con ojo crítico

Trata la respuesta de la IA como la opinión de un amigo, no como el diagnóstico de un médico. Lee cada señal de alerta que mencione y pregúntate: ¿realmente coincide con el mensaje que recibí? Muchas estafas se apoyan en la urgencia ("actúa en las próximas 24 horas"), se hacen pasar por una marca de confianza, incluyen un enlace que casi parece correcto, te piden hacer clic o iniciar sesión, o exigen pago en tarjetas regalo o criptomonedas. Si la IA señala algo que en realidad no está en tu mensaje, ignora ese punto: a veces los asistentes se equivocan, y tú sigues siendo el experto en tu propia bandeja de entrada. Al revés, si algo te huele mal pero la IA no lo

ha mencionado, confía en tu instinto y pasa al siguiente paso. Desconfía especialmente si la IA dice "esto parece correcto" pero el mensaje te pide hacer clic en un enlace, enviar dinero o compartir un código: ese es el momento de buscar una segunda opinión humana. 💬 Ejemplo: si la IA dice "el dominio del remitente no coincide con la empresa", compruébalo tú mismo: ¿la dirección "de" realmente termina en la web oficial de la empresa, o en algo como "soporte-servicio.example.com"? Sabrás que funcionó cuando puedas repetir, con tus propias palabras, dos o tres motivos concretos que la IA dio para su veredicto.

Paso 5 — Verifica tú mismo los enlaces y nombres de remitente

No hagas clic en nada todavía. Si el mensaje contiene un enlace, pasa el ratón por encima en el ordenador o mantén pulsado en el móvil: verás el destino real. Las empresas reales usan su propio dominio corto (la página oficial de la marca), mientras que los enlaces de estafa suelen esconderse tras palabras como "acceso-seguro" o "verificar-cuenta" en una dirección totalmente distinta. La regla es simple: si el enlace no termina en la web real de la marca, casi con total seguridad es falso. Si el mensaje dice ser de una persona que conoces, contacta con ella por un canal que ya consideres de confianza — una llamada o un chat aparte — y nunca respondas al mensaje sospechoso. Si dice ser de una empresa como tu banco, busca el número oficial en su web real (escribe tú la dirección, no uses ningún número que aparezca en el mensaje) y pregunta directamente. 💬 Ejemplo: un enlace de seguimiento de "USPS" apunta a `usps-tracking.example.com/pagar-ahora` — eso NO es usps.com, así que es falso. Sabrás que funcionado cuando hayas confirmado que el mensaje es genuino por un canal de confianza, o hayas llegado a la conclusión de que es una estafa.

Paso 6 — Decide qué hacer y no actúes con prisas

Ahora elige una de tres opciones. Si es una estafa: bórrala, bloquea al remitente y, si se hacía pasar por una empresa, repórtala a esa empresa a través de su web oficial. Si es genuina pero inesperada: aun así no hagas clic en los enlaces que contiene; entra tú directamente en la web real de la empresa. Si no estás seguro: déjala una hora y vuélvela a leer en frío; la urgencia es la herramienta favorita del estafador, y una segunda mirada tranquila suele revelar más. Cuéntaselo a un amigo o familiar si estás preocupado: un segundo par de ojos humanos sigue pillando cosas que la IA se pierde, sobre todo en estafas por voz o vídeo. La idea de usar una IA aquí es frenar, no precipitarse. 💬 Ejemplo: manda a un familiar — "Oye, acabo de recibir un mensaje raro sobre mi cuenta de Amazon, ¿te suena legítimo? [pega]" Sabrás que funcionó cuando tomes una acción concreta en los próximos minutos — borrar, reportar o verificar por un canal de confianza — en lugar de dejar el mensaje sin leer en tu bandeja.

Errores comunes

- **Pegar contraseñas reales o números completos de tarjeta.** La IA no los necesita, y guardarlos en un chat genera riesgo si tu cuenta se ve comprometida. Solución: tacha antes (paso 2), o describe el mensaje con tus propias palabras en lugar de pegarlo.
- **Confiar en el veredicto de la IA sin comprobarlo.** A veces los asistentes se saltan estafas evidentes o marcan mensajes limpios por error. Solución: contrasta siempre las señales concretas que menciona con el mensaje real (paso 4) y verifica con la empresa de verdad (paso 5).
- **Hacer clic en el enlace sospechoso "solo por ver qué pasa".** Es justo lo que el estafador quiere. Solución: pasa el ratón o mantén pulsado para leer la URL real, y nunca inicies sesión a través de un enlace dentro de un mensaje inesperado; entra tú en la web de la empresa.

Qué significa esto para ti

- **En el día a día:** la próxima vez que te llegue un SMS tipo "tu paquete no ha podido entregarse" o un correo tipo "tu cuenta será bloqueada", podrás obtener una segunda opinión clara en menos de un minuto en vez de adivinar si pulsar o no.
- **Para tu negocio o trabajo:** si gestionas facturas o pagos a proveedores, esta misma comprobación de 5 minutos ante solicitudes de cambio de datos bancarios puede frenar una estafa de correo corporativo (en la que un criminal se hace pasar por un proveedor o un jefe para desviar un pago) — una de las estafas más caras que hay.
- **Si estás empezando:** elige el asistente de IA que ya uses, guarda esta guía en favoritos y pruébala la próxima vez que algo te parezca aunque sea ligeramente raro. Practica primero con un correo basura si quieres ganar confianza.

Pruébalo ahora

Busca un mensaje de tu bandeja del que no estés seguro, tacha los datos personales, pégalo en tu asistente de IA y pídele que señale las señales de alerta. Cinco minutos, una conversación nueva, una decisión más segura.

Cómo detectar hoy una llamada, un correo o una voz falsa generados con IA

6 pasos · Léela en la web

> **En resumen:** las estafas con IA ya suenan como tu familia, escriben como una persona real y falsifican el número de tu banco. Una pausa de cinco segundos más una palabra secreta familiar atrapa a casi todas: así lo montas hoy.

Cuando termines esta guía sabrás reconocer las tres estafas con IA más comunes que circulan ahora, qué comprobación de un segundo frena cada una y qué hacer en el momento en que te das cuenta de que algo va mal. Está escrita para cualquiera que use un teléfono y el correo — no hace falta tener conocimientos técnicos y aquí no hay que comprar ni instalar nada.

Antes de empezar

- Un teléfono (móvil o fijo) y una cuenta de correo que consultes con regularidad.
- Unos 10 minutos en total, sobre todo para configurar una protección que te servirá durante años.
- Una o dos personas de confianza — un familiar o un amigo cercano — para el paso de la palabra secreta familiar.

Paso 1 — Detecta la señal en una llamada de "emergencia familiar"

Cuando alguien llame haciéndose pasar por un familiar con un problema, lo primero que debes hacer es **colgar** y marcar tú el número de esa persona — el que ya tienes guardado en tu teléfono. Lo habitual es ver en pantalla un nombre conocido en la identificación de llamadas, a veces con una voz desconocida que pregunta "¿a que no sabes quién soy?" para que seas tú quien complete el hueco. Después de colgar y volver a marcar, un familiar real contestará o te devolverá la llamada; una línea de estafa a menudo se corta, reproduce un buzón genérico o directamente no suena. Si tu identificador no muestra nada o un número aleatorio, se aplica la misma regla: nunca sigas la llamada. La clonación de voz (una IA que copia la voz de alguien con unos segundos de audio) ya es tan buena que hasta hijos adultos de la "persona que llama" no notan la diferencia en los primeros segundos. Ejemplo: "Abuela, soy yo, he tenido un accidente, por favor mándame dinero" → cuelga → marca el número guardado de tu nieto. Sabrás que ha funcionado cuando la persona real conteste a tu segunda llamada y confirme que no te ha telefoneado.

Paso 2 — Descubre a un falso agente de atención al cliente

Cada vez que alguien te llame A TI sobre un problema con tu cuenta — banco, teléfono, mensajería, internet — cuelga y devuelve la llamada a la empresa usando el número impreso en tu último extracto o el que aparece en la web oficial de la compañía. Lo habitual en la llamada original es un "agente" que ya sabe tu nombre, los últimos cuatro dígitos de tu tarjeta o transacciones recientes; la voz suena profesional y a menudo se oye ruido de oficina de fondo para parecer legítimo. Una empresa real te dejará colgar y llamar de nuevo; un estafador te presionará para que "sigas en la línea" o se negará a darte un número que puedas verificar. Si la persona que llama insiste en que no puede esperar tu rellamada porque el caso es "urgente", esa es en realidad la mayor señal de alarma. 💬 Ejemplo: "Hola, soy María, del departamento de fraude de tu banco" → cuelga → busca el número del banco en tu extracto → márcalo tú. Sabrás que ha funcionado cuando la locución grabada de la empresa real responda y la persona a la que llegues no tenga constancia de la llamada anterior.

Paso 3 — Identifica el correo de phishing escrito por IA

Antes de hacer clic en cualquier enlace de un mensaje que te pide actuar, **pasa el ratón por encima del remitente y de cualquier enlace** para ver a dónde apuntan realmente. Verás un correo que parece venir de una empresa de mensajería, una plataforma de streaming o tu banco — el texto es fluido, amable y sin las erratas que tenían las estafas de antes; a menudo incluye una pequeña urgencia tipo "actúa en 24 horas" para empujarte a dejar de pensar. Al pasar el ratón verás un dominio distinto — algo como `usps-tracking.example.com/pay-now` en lugar de una dirección postal real, o una cadena larga de letras que parecen aleatorias. En el móvil puedes mantener pulsado un enlace para previsualizarlo; si no ves claramente a dónde va, no lo toques. El phishing escrito por IA es convincente porque la gramática ahora es perfecta — la señal ya no está en la ortografía, está en el destino. 💬 Ejemplo: "Tu paquete está retenido en aduanas, paga una pequeña tasa para liberarlo" → pasa el ratón por el enlace → ves que apunta a un dominio que no conoces → elimínalo. Sabrás que ha funcionado cuando sepas decir el destino real del enlace antes de decidir si haces clic.

Paso 4 — Usa la comprobación de un segundo que sirve para las tres

Cuando cualquier mensaje o llamada resulte urgente, asustadizo o demasiado bonito para ser verdad, haz una pausa de cinco segundos y plantéate una sola pregunta: **"¿Tendría un problema si no hago nada en los próximos diez minutos?"**. Lo que notarás en esos cinco segundos son tus propias manos queriendo actuar rápido — y la llamada o el correo empujándote a "hacerlo ya". La mayoría de estafas se desinflan con una pausa de diez minutos: el "agente" cuelga y pasa al siguiente, el correo puede quedarse sin leer hasta mañana, y tu "nieto" sigue sano esté donde esté. Algunas estafas añaden un contador falso para combatir justo este reflejo; recuerda que una emergencia real de una empresa real seguirá siendo una

emergencia real dentro de diez minutos, y un ser querido real entenderá una breve espera. 💬 Ejemplo: "Paga esta tasa ahora o tu cuenta se cierra en cinco minutos" → pausa → haz la pregunta de los diez minutos → casi siempre la respuesta es no. Sabrás que ha funcionado cuando la urgencia se deshaga y tomes tu siguiente decisión desde la calma en lugar del pánico.

Paso 5 — Crea hoy mismo una palabra secreta familiar

Elige una palabra poco habitual con una o dos personas de confianza y quedaos en que cualquier llamada de "emergencia" tendrá que incluir esa palabra antes de que se mueva dinero. Basta con un breve chat de grupo, un mensaje de texto o incluso una conversación en persona — sin apps ni gestores de contraseñas. Esto te da una prueba privada que ni la mejor clonación de voz puede superar: el estafador simplemente no conoce tu palabra, y tu familiar real sí. Si tu familia está repartida, acordad en su lugar una pregunta privada con una respuesta privada — "¿cómo se llamaba el gato en casa de la abuela?" funciona mejor que una sola palabra cuando hay niños por medio. Apunta la palabra una vez, compártela con quien la necesite y no la publiques en ningún sitio de internet. 💬 Ejemplo: Manda un mensaje a tu hermana: "Si alguien llama diciendo que soy yo y pide dinero, pídele nuestra palabra secreta. La palabra es ornitorrinco." Sabrás que ha funcionado cuando al menos dos personas de tu círculo puedan decir la palabra de memoria sin tener que mirarla.

Paso 6 — Sabe qué hacer en el momento en que te das cuenta de que es una estafa

Cuelga, cierra el correo o deja de responder — y luego **denúncialo**, aunque no hayas perdido ni un céntimo. Te seguirá latiendo el corazón deprisa, el "agente" ya no estará y el correo seguirá en tu bandeja; es normal. Denunciar ayuda a proteger a la siguiente persona, y puedes reenviar correos de phishing a reportphishing@apwg.org y presentar una denuncia en reportfraud.ftc.gov (EE. UU.) o en el portal de ciberdelitos de tu país. Si SÍ has enviado dinero o compartido un código de un solo uso, llama de inmediato al número de fraude de tu banco — el de la parte trasera de tu tarjeta — y pide una devolución; aquí la velocidad importa más que la vergüenza. 💬 Ejemplo: Te das cuenta a mitad de la llamada de que el "agente del banco" era falso → cuelgo → llamo al número de mi tarjeta de débito → cuento lo ocurrido → pido que vigilen la cuenta. Sabrás que ha funcionado cuando la llamada haya terminado, se lo hayas contado a un canal oficial y tus cuentas y contraseñas sigan sin cambios.

Errores habituales

- **Fiarte del identificador de llamadas.** Los estafadores pueden falsificar cualquier nombre o número. Que en la pantalla salga "Hija" o "Tu banco" no prueba nada — cuelga y vuelve a marcar con un número que hayas guardado tú.

- **Ceder ante la "cuenta atrás de diez minutos".** Ese temporizador forma parte de la estafa. Las empresas reales y las emergencias familiares reales no desaparecen porque respires un momento; pausar siempre es seguro.
- **Compartir un código de un solo uso "para verificarte".** Ninguna empresa legítima te pedirá jamás un código que ella misma te acaba de enviar. En el instante en que alguien te pida esa cadena de números, la conversación es una estafa — sin discusión.
- **Saltarte la palabra secreta familiar porque "a nosotros no nos va a pasar".** Configurarla lleva un minuto hoy y puede ahorrarte una cantidad de dinero que cambie tu vida. Trátala como un detector de humo: aburrido de instalar, impagable cuando suena.

Qué significa esto para ti

- **En el día a día:** la próxima vez que tu teléfono suene con una voz angustiada, sabrás colgar primero y devolver la llamada a un número de confianza — en lugar de enviar dinero en los primeros treinta segundos.
- **En tu negocio o trabajo:** cualquier equipo que gestione facturas, correos de proveedores o transferencias bancarias debería adoptar una palabra secreta igual que las familias; las estafas con voz clonada ya han golpeado a departamentos de finanzas y de RR. HH., no solo a las abuelas.
- **Si estás empezando:** haz una sola cosa hoy — manda a un familiar la palabra secreta del paso 5. Ese único mensaje te protege más que cualquier aplicación antivirus de tu teléfono.

Pruébalo ahora

Ahora mismo, abre tu app de mensajería y manda a una persona de confianza: *"Si alguna vez te llamo desesperado y te pido dinero o códigos, pídeme nuestra palabra secreta. La palabra es [elige algo raro, como 'ornitorrinco' o 'cuchara-de-plata']."* Ese mensaje lleva menos de dos minutos y es la defensa más valiosa de toda esta guía.

Desactiva el historial y el guardado de datos en ChatGPT, Claude y Gemini

6 pasos · Léela en la web

> **En resumen:** Puedes impedir que ChatGPT, Claude y Gemini usen tus conversaciones para entrenar futuros modelos de IA con solo activar un interruptor de privacidad en los ajustes de cada servicio, aunque los nombres exactos de los menús cambian con frecuencia y, aun desactivando el entrenamiento, no se eliminan las conversaciones que esas empresas ya hayan guardado.

Por defecto, ChatGPT, Claude y Gemini guardan tus chats y pueden usarlos para entrenar futuras versiones de sí mismos. Esta guía te muestra cómo desactivarlo en cada servicio, tanto en la web como en el móvil, y explica qué cambia realmente con cada interruptor, y qué cosas, en silencio, no cambia.

Antes de empezar

- Una cuenta **personal** gratuita o de pago en cada servicio de IA que quieras ajustar. Las cuentas de trabajo o de colegio suelen estar gestionadas por tu empresa y quizá no te permitan cambiar estos ajustes.
- Lo más fácil es hacerlo desde un navegador web. Las apps de móvil también sirven, pero los menús cambian de sitio con más frecuencia.
- Reserva unos **15 minutos en total**, más o menos cinco minutos por servicio.
- Asegúrate de haber iniciado sesión en cada servicio antes de empezar, para que los ajustes se guarden de verdad.

Paso 1 — Abre los controles de datos de ChatGPT

Entra en **chatgpt.com** desde un navegador e inicia sesión si todavía no lo has hecho. Haz clic en tu **foto de perfil o tus iniciales** en la esquina superior derecha; aparecerá un pequeño menú. Busca una opción como **Settings** (a veces aparece como un icono de engranaje ⚙) y haz clic. En el panel de ajustes que se abre a la izquierda de la página, busca una sección llamada **Data Controls** y haz clic. Si tu pantalla se ve distinta, prueba con una sección llamada **Privacy**; las versiones antiguas usaban ese nombre.

Si estás en el móvil, toca las **dos líneas o los tres puntos** en la esquina superior de la app de ChatGPT y, después, toca tu nombre en la parte superior del menú que aparece. Deberías ver el mismo camino Settings → Data Controls.

Sabrás que lo has hecho bien cuando veas una lista de interruptores: busca los que mencionen "Improve", "training" o "history".

Paso 2 — Desactiva el interruptor de entrenamiento de ChatGPT

En la página de Data Controls, busca el interruptor con una etiqueta parecida a **"Improve model for everyone"** o **"Chat history & training"**, ya que el nombre exacto cambia a menudo, así que lee la descripción de cada interruptor. Haz clic en el interruptor para que pase a la posición de **apagado** (suele volverse gris o perder su color). Puede que aparezca una pequeña ventana pidiéndote que confirmes: haz clic en **Confirm** o **Save**. Es posible que la página se refresque o que tengas que cerrarla y volver a abrirla para que se vea el cambio.

Si no ves ningún interruptor de entrenamiento, puede que tu cuenta esté en un plan o en una región donde la opción está oculta. En ese caso, sal de la página y vuelve a comprobarla en unas semanas: los ajustes se van añadiendo con el tiempo.

Sabrás que ha funcionado cuando el interruptor siga en gris (apagado) tras refrescar la página y ya no muestre ningún indicador de "encendido". Incluso después de desactivarlo, OpenAI puede seguir guardando tus chats más recientes hasta 30 días por motivos de seguridad, así que evita pegar nada realmente sensible mientras tanto.

Paso 3 — Abre los ajustes de privacidad de Claude

Entra en **claude.ai** e inicia sesión. Haz clic en tu **foto de perfil o tus iniciales**, normalmente abajo a la izquierda en la web o arriba a la derecha en el móvil. En el menú que aparece, elige **Settings**. Dentro de Settings, busca una pestaña o sección llamada **Privacy**, **Account** o **Profile**; el nombre exacto varía. Dentro de esa sección, busca un interruptor que mencione **"help improve Claude"** o **"use my chats for training"**.

En la app móvil de Claude, el mismo camino suele funcionar: toca tu **icono de perfil**, luego **Settings** y después busca la sección de privacidad. Si la app se abre directamente en un chat, busca un **icono de engranaje** ⚙ en la esquina superior.

Sabrás que lo has hecho bien cuando veas una sección de privacidad con al menos un interruptor de encendido/apagado relacionado con entrenar o mejorar al asistente.

Paso 4 — Desactiva el interruptor de entrenamiento de Claude

Haz clic en el interruptor junto a **"Help improve Claude"** (o la frase similar que aparezca en tu pantalla) para que pase a la posición de **apagado**. Algunas personas ven una ventana de confirmación: acéptala. El cambio suele guardarse al instante sin un botón de Save aparte, pero si lo ves, haz clic en él.

Unas notas honestas aquí. En algunas cuentas gratuitas de Claude, la opción de no participar en el entrenamiento puede ser limitada o no estar disponible; Anthropic ha actualizado esta política más de una vez. Si no ves el interruptor, comprueba si has iniciado sesión con un **correo de trabajo o del colegio**: esas cuentas las gestiona un administrador y no se pueden cambiar desde aquí. Si pagas Claude, puede que veas el interruptor en un sitio algo diferente, en la misma zona de Settings, pero quizá dentro de una sub-pestaña de **Privacy**.

Sabrás que ha funcionado cuando el interruptor esté gris y siga gris al refrescar la página. Si el ajuste no parece quedarse guardado, cierra sesión, vuelve a iniciarla y comprueba de nuevo: a veces tarda un momento en sincronizarse.

Paso 5 — Abre los controles de actividad de Gemini

Gemini lo gestiona de forma un poco diferente porque vive dentro de tu cuenta de Google. Entra en **gemini.google.com** o abre la app de Gemini. Haz clic en tu **foto de perfil** en la esquina superior. Busca **Settings** o **Activity**: en la web puede que veas un enlace directo llamado **"Gemini Apps Activity"** que te lleva directamente a los controles de tu cuenta de Google. Haz clic en él.

En el móvil, toca tu **foto de perfil** en la esquina superior de la app de Gemini y busca **Gemini Apps Activity** en el menú que aparece. Si no está, puede que tu versión de la app la haya movido: prueba a buscar **"Gemini"** en la app de ajustes de tu teléfono o abre **myactivity.google.com** en un navegador.

Sabrás que lo has hecho bien cuando se abra una página de Google con un interruptor llamado **"Gemini Apps Activity"** y una breve descripción de lo que hace.

Paso 6 — Desactiva Gemini Apps Activity

En la página de Gemini Apps Activity, haz clic en el interruptor de arriba para que pase a **off**. Puede que aparezca un pequeño panel de confirmación explicando qué va a cambiar: léelo y luego confirma. También puede que veas una opción aparte para **eliminar la actividad pasada**; si quieres empezar de cero, haz clic también ahí y sigue los pasos.

Este único interruptor de Google cubre la app web de Gemini, la app móvil de Gemini y cualquier otro sitio donde Gemini guarde tus chats en tu cuenta de Google. Desactivarlo no borra tus conversaciones pasadas automáticamente: tienes que hacerlo por separado si quieres que desaparezcan. Y recuerda: desactivar Gemini Apps Activity no afecta al historial de Google Search ni al de YouTube, que se controlan con interruptores diferentes.

Sabrás que ha funcionado cuando el interruptor esté apagado y una línea de texto confirme que la actividad está en pausa. Puedes comprobarlo iniciando un nuevo chat y volviendo después a

esta página: la actividad no debería aparecer ahí.

Errores frecuentes

- **Buscar un botón de "borrar mis datos".** Ninguno de estos interruptores borra lo que la empresa ya tiene guardado. Si quieres que se eliminen chats antiguos, cada servicio tiene una opción aparte de **Delete** o **Clear conversations**, e incluso así, las empresas suelen guardar copias durante un tiempo por motivos legales o de seguridad.
- **Pensar que un solo interruptor lo arregla todo.** En Gemini sobre todo hay **varios** controles de actividad de Google. Desactivar Gemini Apps Activity no afecta al historial de Search, al de YouTube ni a otros productos de Google. Comprueba cada uno por separado si te importan todos.
- **Confiar en el interruptor para siempre.** Las empresas actualizan sus ajustes y políticas con regularidad. La opción de exclusión que activas hoy puede desaparecer en silencio o cambiar de significado en seis meses: ponte un recordatorio en el calendario cada pocos meses para volver a comprobarlo.

Qué significa esto para ti

- **En el día a día:** desactivar esto es un buen hábito si sueles pegar algo personal en un chat de IA, como nombres, direcciones, preguntas de salud o correos de trabajo. No te vuelve invisible, pero impide que tus palabras se usen para enseñar al próximo modelo.
- **Para tu negocio o tu trabajo:** si manejas datos de clientes, contratos o cualquier cosa confidencial, no te apoyes solo en estos interruptores. Muchas empresas dicen claramente que las exclusiones no cubren solicitudes legales ni la detección interna de abusos. Usa una cuenta de IA aparte con datos sintéticos o quédate con planes enterprise que prometan mayor privacidad.
- **Si estás empezando:** elige el servicio que más uses, abre sus ajustes y apaga ahora mismo el interruptor de entrenamiento. Ese único gesto es lo más útil que puedes hacer hoy; los otros dos puedes dejarlos para luego.

Pruébalo ahora

Abre el asistente de IA que hayas usado más recientemente, haz clic en tu foto de perfil, busca los ajustes de privacidad o de datos y apaga el interruptor de entrenamiento. Tardarás más o menos un minuto y verás exactamente cómo están tus ajustes hoy mismo.

Lo que nunca deberías compartir con un asistente de IA: guía de privacidad

6 pasos · Léela en la web

> **En resumen:** Una guía sencilla para que sepas qué datos personales jamás deberías pasarle a un asistente de IA y cómo borrar cualquier cosa sensible que ya le hayas enseñado sin querer.

Cuando termines esta guía vas a saber cuáles son los tres grupos de datos personales que deben quedarse siempre fuera de los chats, cómo detectarlos en tu historial y qué hacer si ya se los has enseñado a una IA. Está pensada para cualquier persona que habla con herramientas como ChatGPT, Gemini, Claude o Copilot.

Antes de empezar

- Un dispositivo con conexión a internet (móvil, tablet u ordenador).
- Una cuenta en al menos un asistente de IA (por ejemplo, ChatGPT, Gemini o Claude: con el plan gratuito basta).
- Unos 5 minutos sin interrupciones.

Paso 1 — Identifica los tres grupos de datos privados

Abre el asistente de IA que uses más a menudo y empieza una conversación nueva. En pantalla verás un cuadro de texto vacío esperando tu pregunta y un botón “Enviar” a la derecha. Cuando escribas, el asistente empezará a generar una respuesta debajo de tu mensaje. Si el diseño se ve diferente, busca un campo que diga “Pregunta lo que quieras” o algo parecido. Ejemplo: “¿Qué tipos de información personal no debería compartir contigo?”. Sabrás que ha funcionado cuando el asistente te liste tres grupos: (1) datos identificativos (nombre, dirección, números de documento), (2) datos financieros (cuentas bancarias, números de tarjeta) y (3) datos de salud o biométricos.

Paso 2 — Revisa tus chats recientes buscando esos grupos

Abre la pantalla del historial de chats, normalmente entrando desde un icono de reloj o un menú llamado “Historial”. Verás una lista de conversaciones pasadas, cada una resumida con su primera línea. Al elegir una conversación se abre el intercambio completo, donde puedes desplazarte por todos los mensajes que enviaste. Si no ves un icono de reloj, busca un menú de tres líneas y elige “Chats recientes”. Ejemplo: escribe “Muéstrame mis últimos cinco mensajes” (muchos asistentes admiten una petición sencilla como esta). Sabrás que ha funcionado cuando el asistente muestre los prompts recientes que escribiste, para que puedas revisarlos.

Paso 3 — Borra o corrige cualquier mensaje con datos privados

En la conversación abierta, localiza el mensaje que tenga un dato identificativo, financiero o de salud. La mayoría de asistentes te dejan mantener pulsado el mensaje para ver opciones como “Borrar” o “Eliminar”. Elige “Borrar” y confirma cuando aparezca el aviso “¿Estás seguro?”. Si la opción no está a la vista, busca un icono de papelera o un botón “Limpiar” en la parte inferior del chat. 💬 Ejemplo: después de seleccionar el mensaje, escribe “Borra este mensaje”. Sabrás que ha funcionado cuando el mensaje desaparezca de la vista del chat y el historial ya no muestre ese contenido.

Paso 4 — Desactiva el guardado del historial (si tu plan lo permite)

Abre el panel de ajustes, normalmente tocando tu foto de perfil o un icono en forma de engranaje en la esquina superior derecha. La pantalla de ajustes mostrará interruptores como “Guardar historial de chat” o “Almacenamiento de conversaciones”. Apaga ese interruptor; puede que aparezca un aviso diciendo que el cambio solo se aplica a los chats futuros. En el plan gratuito, algunos asistentes solo dejan “borrar el historial” en vez de desactivarlo para siempre, así que busca un botón que diga “Borrar todos los chats”. 💬 Ejemplo: “Desactiva el historial de chat”. Sabrás que ha funcionado cuando el interruptor cambie de color o pase a la posición “off” y la app confirme “Tus chats no se guardarán”.

Paso 5 — Comprueba que no quede nada sensible guardado

Empieza una conversación nueva y escribe una pregunta inofensiva como “¿Cuál es la capital de Estados Unidos?”. El asistente responderá debajo del mensaje. Luego abre el historial otra vez y verifica que solo aparezca esta pregunta nueva y nada personal. Si todavía ves mensajes antiguos con datos privados, repite los pasos 2 a 4 con los que queden. 💬 Ejemplo: “Muéstrame mis chats recientes”. Sabrás que ha funcionado cuando la lista solo contenga la pregunta inofensiva que acabas de hacer.

Paso 6 — Pide una eliminación completa de datos si crees que algo se ha escapado

Si sospechas que compartiste algo que el asistente no puede borrar por sí mismo (por ejemplo, un texto enterrado en una conversación muy larga), busca un enlace de “Ayuda” o “Soporte” en el menú de ajustes. La mayoría de proveedores tienen un formulario de contacto o un correo llamado “Solicitud de privacidad de datos”. Envía un mensaje breve que diga: “Por favor, eliminen todos los datos personales asociados a mi cuenta [tu correo]”. En algunas plataformas quizá tengas que verificar antes tu identidad. 💬 Ejemplo: “Estimado soporte: por favor,

eliminen cualquier dato identificativo de mi cuenta". Sabrás que ha funcionado cuando recibas un correo de confirmación o un mensaje del sistema diciendo "Hemos recibido tu solicitud".

Errores habituales

- **Compartir direcciones completas o números de documento:** si copias y pegas la dirección de tu casa, el asistente ya conoce una ubicación que quizá no quieras que quede guardada. Solución: borra el mensaje y desactiva el historial como se explica en los pasos 3 y 4.
- **Pegar números de cuenta o de tarjeta:** incluso un número parcial puede usarse para fraude. Borra el mensaje de inmediato y avisa a tu banco para que vigile cualquier movimiento raro.
- **Asumir que el asistente borra los datos por ti:** borrar un mensaje solo lo quita de tu vista; el proveedor puede seguir guardando una copia en sus servidores. Usa siempre la opción integrada "Borrar todos los chats" o envía una solicitud formal de eliminación (paso 6).

Qué significa esto para ti

- **En el día a día:** puedes pedirle a una IA que te ayude con la lista de la compra sin miedo a que tu número de teléfono o tu dirección queden guardados para siempre.
- **En tu negocio o trabajo:** los números confidenciales de clientes y las cifras de contratos se quedan fuera de los datos de entrenamiento de la IA, lo que reduce el riesgo de filtraciones accidentales.
- **Si estás empezando:** el primer movimiento más fácil es desactivar el guardado del historial: es un solo interruptor que protege todas las conversaciones futuras.

Pruébalo ahora mismo

Abre el asistente de IA que más uses, busca el engranaje de ajustes y apaga la opción "Guardar historial de chat". En menos de dos minutos habrás dado un paso firme para mantener tu información privada fuera de los archivos generados por la IA.

Cómo detectar un correo de phishing o un mensaje fraudulento antes de hacer clic

6 pasos · Léela en la web

> **En resumen:** Copia cualquier correo sospechoso o mensaje fraudulento en un asistente de IA con un breve aviso para detectar fraudes. Te mostrará las señales de alerta —falsa urgencia, direcciones del remitente que no coinciden con la marca y enlaces dudosos— para que decidas si debes borrar, bloquear o denunciar el mensaje antes de hacer clic en nada.

Ese presentimiento de «espera, algo no va bien» que tienes cuando llega un correo o mensaje extraño puede confirmarlo la IA —o tranquilizarte— en unos 30 segundos. Esta guía te muestra exactamente qué escribir y cómo leer la respuesta para no pasar por alto las señales de advertencia. Funciona con ChatGPT, Gemini, Claude, Copilot, Le Chat, DeepSeek o cualquier otro asistente que ya uses. Una aclaración importante: la IA puede detectar muchas cosas, pero no puede hacer clic en el enlace por ti ni verificar si el remitente es real; la decisión final sigue siendo tuya.

Antes de empezar

- Una cuenta en cualquier asistente de IA que ya uses (las versiones gratuitas sirven perfectamente; ChatGPT, Gemini, Claude, Microsoft Copilot, Le Chat y DeepSeek pueden hacerlo).
- El mensaje sospechoso, abierto en tu pantalla, pero sin tocar ningún enlace ni enviar respuestas.
- Unos 2 minutos.

Paso 1: Lee el mensaje con calma antes de hacer nada

Abre el correo o mensaje sospechoso y haz exactamente una cosa: léelo de principio a fin sin tocar el ratón. Fíjate en tres elementos: un remitente o número que no reconoces, un tono que te presiona para actuar AHORA y cualquier enlace inesperado. Verás el mensaje exactamente como lo harías normalmente; todavía no cambia nada. 💬 Ejemplo: si dice «Tu paquete se ha retrasado, confirma aquí: bit.ly/x9zq», limítate a leer las palabras. No toques el enlace. Si se ve diferente: en algunas aplicaciones de mensajería, mantener pulsado un enlace muestra una pequeña vista previa de la URL; eso puede ser útil, pero no lo abras. El objetivo aquí es leer. Sabrás que lo has hecho bien cuando puedas describir el mensaje en una sola frase sin haber hecho clic en nada.

Paso 2: Abre cualquier asistente de IA que ya uses

En tu teléfono u ordenador, abre la aplicación o el sitio web del asistente de IA en el que ya tengas una cuenta. No necesitas ninguna «herramienta contra el fraude» especial: cualquier asistente de uso general puede hacer este trabajo. Verás un cuadro de chat vacío, normalmente con una etiqueta como «Mensaje», «Pregunta lo que quieras» o algo parecido. Si se ve diferente: cada asistente tiene su propia pantalla inicial, pero el cuadro de chat casi siempre está en el centro de la pantalla en el ordenador o en la parte inferior del móvil. Si no sabes qué hacer, inicia un «Chat nuevo». Sabrás que lo has hecho bien cuando veas un cursor parpadeando y un botón de envío, listos para que escribas.

Paso 3: Pega primero un aviso para «detectar fraudes»

Antes de pegar el mensaje sospechoso, dale a la IA una instrucción de una sola línea para definir su función. Escríbela en el cuadro de chat y pulsa enviar. La IA responderá con una breve confirmación —normalmente algo como «Claro, pégalo aquí»— que la prepara para centrarse en patrones de fraude en lugar de dar consejos genéricos. Esta única línea convierte un chat normal en una comprobación de fraude. 💬 Ejemplo: «Actúa como detector de fraudes. Lee el mensaje que voy a pegar y enumera todas las señales de alerta que puedas ver. Después, dime claramente si es seguro responder, hacer clic en algún enlace o abrir algún archivo adjunto. Si no estás seguro, dilo». Si se ve diferente: algunos asistentes responden con un párrafo de configuración más largo; no pasa nada. Lo importante es que ya sabe que debe centrarse en las estafas. Sabrás que lo has hecho bien cuando la IA responda algo como «Entendido: pega el mensaje» o te haga una pregunta aclaratoria.

Paso 4: Pega el mensaje sospechoso completo y envíalo

Selecciona todo el mensaje —nombre del remitente, asunto, contenido y texto del enlace—, pero elimina tu propio nombre, dirección o cualquier número de cuenta que aparezca, ya que la IA no los necesita para detectar problemas. Pégalo en el cuadro de chat y pulsa enviar. La IA ya tiene el texto original y comienza a analizarlo. Si se ve diferente: si el mensaje es especialmente largo —por ejemplo, una factura falsa—, puedes pegar una versión recortada, pero conserva exactamente la frase inicial y todos los enlaces tal como aparecen, porque ahí suelen estar las señales de alerta. 💬 Ejemplo de lo que puedes pegar: «De: Soporte de Amazon <> — Estimado cliente: tu cuenta ha sido suspendida debido a una actividad inusual. Haz clic aquí para verificarla: hhtp://amazon-secure.example.com/login — responde en un plazo de 24 horas». Sabrás que lo has hecho bien cuando la IA empiece a responder con una lista de elementos concretos que ha detectado, normalmente entre 4 y 8.

Paso 5: Lee las señales de alerta y confía en las evidentes

La respuesta de la IA suele ser una lista de comprobación: cosas como «el dominio del remitente no coincide con Amazon», «la URL utiliza un cero en lugar de una “o”», «el mensaje crea

urgencia con un plazo de 24 horas» o «el saludo es genérico y no incluye tu nombre». Lee cada punto y pregúntate: «¿Coincide con lo que estoy viendo en mi pantalla?». Empezarás a reconocer el patrón la próxima vez sin ayuda. 💬 Ejemplo de respuesta de la IA: «Señales de alerta: el dominio del remitente es amazon-support.example.com, no amazon.com; el enlace empieza por “hhtp” (error tipográfico) y utiliza un cero en “amazon”; táctica de urgencia (“en un plazo de 24 horas”); saludo genérico. Veredicto: NO hagas clic ni respondas. Denuncia el mensaje y bórralo». Si se ve diferente: la IA puede usar expresiones más suaves («probablemente es un intento de phishing», «parece sospechoso»). El significado es el mismo. Si alguna vez dice «no estoy seguro», tómalo en serio y considera el mensaje una estafa. Sabrás que lo has hecho bien cuando puedas mencionar al menos tres razones concretas por las que el mensaje es sospechoso y explicarlas en voz alta.

Paso 6: Decide, y después denuncia y borra el mensaje

Cuando la IA te dé su veredicto, decide según la importancia de las señales de alerta. Si aparece aunque sea una señal importante —un dominio dudoso del remitente, presión para actuar AHORA o una solicitud de contraseñas o pagos—, no hagas clic ni respondas. Usa la opción integrada «Denunciar phishing» o «Denunciar como spam» de tu aplicación de correo o mensajería, si la encuentras, y después borra el mensaje. El mensaje peligroso desaparecerá de tu pantalla y quedará señalado ante la plataforma. 💬 Ejemplo de una buena pregunta de seguimiento para la IA: «¿Puedes explicarme en una frase qué significa una “táctica de urgencia”, para saber qué buscar la próxima vez?». Si se ve diferente: Gmail, Outlook, Apple Mail y la mayoría de los operadores móviles tienen un botón para denunciar, normalmente dentro de un menú de tres puntos «⋯» o cerca de «Marcar como» / «Mover a». Si no encuentras la opción para denunciar, borrarlo es suficiente: lo importante es que no respondas. Sabrás que lo has hecho bien cuando el mensaje ya no esté en tu bandeja de entrada, lo hayas señalado ante tu servicio de correo u operador y no hayas tocado el enlace.

Errores frecuentes

- Responder al estafador solo para preguntar «¿esto es real?» —responder confirma que tu dirección está activa y hace que lleguen más mensajes fraudulentos. En su lugar, borra y denuncia el mensaje.
- Confiar en la IA sin leer la explicación —la IA puede pasar cosas por alto. Si el mensaje te «parece» extraño pero la IA dice que está bien, confía en tu intuición y no hagas clic. La IA es una segunda opinión, no la decisión definitiva.
- Pegar datos personales junto con el mensaje —tu nombre, dirección y números de cuenta no son necesarios para detectar señales de alerta. Elimínalos antes de pegar el mensaje, porque todo lo que envías a un chat es procesado por ese servicio.

Qué significa esto para ti

- **En la vida diaria:** Comprobar durante 30 segundos un enlace desconocido antes de hacer clic puede evitarte el robo de contraseñas, el vaciado de tus cuentas bancarias y largas llamadas a tu banco. Se convierte en un hábito que te protege automáticamente.
- **En el trabajo:** Convierte esto en tu primer paso habitual antes de reenviar correos laborales sospechosos al equipo de informática. A menudo detectarás estafas evidentes que el equipo de informática habría señalado igualmente, lo que ahorra tiempo a todos y te gana la reputación de ser la persona más cuidadosa del equipo.
- **Si estás empezando:** Pruébalo con el próximo mensaje extraño que recibas. No tienes que hacerlo todo perfecto; solo necesitas detenerte durante 30 segundos antes de hacer clic.

Pruébalo ahora

La próxima vez que un correo o mensaje te parezca extraño, cópialo en tu asistente de IA con el aviso del paso 3, pega el mensaje completo del paso 4 y lee las señales de alerta del paso 5 antes de tocar nada. Toda la comprobación dura aproximadamente medio minuto.

Cómo detectar estafas de IA, deepfakes y contenido falso

5 pasos · Léela en la web

> **En resumen:** el contenido generado por IA es cada vez más convincente, pero aún puedes protegerte verificando por canales independientes y manteniendo la calma cuando los mensajes generan urgencia.

Al terminar esta guía, sabrás cómo identificar imágenes generadas por IA, voces sintéticas y mensajes engañosos antes de que te engañen. Esta guía práctica está pensada para que cualquier persona navegue por internet de forma segura, sin necesidad de tener un título en informática. Aunque estos consejos son muy útiles, recuerda que la tecnología de IA evoluciona constantemente. Ningún método es 100 % infalible, así que mantenerte al día y evaluar el contenido con sentido crítico es tu mejor defensa. Esta guía te enseña a ser observador, pero no puede decidir por ti sobre la fiabilidad de una fuente.

Antes de empezar

- **Un dispositivo con acceso a internet:** puede ser tu móvil, tableta u ordenador.
- **Entre 10 y 15 minutos de tu tiempo:** suficiente para leer la guía y poner en práctica algunas observaciones.
- **Una mente atenta y tranquila:** la IA cambia rápido, así que la vigilancia constante y la cabeza fría son tus mejores herramientas frente al engaño.
- **No necesitas apps ni cuentas especiales:** nos centraremos únicamente en usar tu capacidad de observación natural y métodos de verificación cotidianos.

Paso 1 — Detecta "fallos" visuales en las fotos

Los **generadores de imágenes con IA** son programas de ordenador que crean imágenes a partir de descripciones de texto, como un pintor digital muy hábil que ha estudiado millones de fotos reales. Aunque son increíblemente buenos, las versiones más antiguas suelen tener problemas con los detalles finos, y examinar las fotos de cerca a veces revela incoherencias.

Cuando veas una foto que te resulte "rara", tu primera acción debería ser ampliar y examinar con detalle ciertos puntos concretos. Al hacer zoom, normalmente verás que la imagen ocupa más espacio en tu pantalla. Una vez ampliada, revisa zonas como manos, pies, accesorios, ropa y fondos en busca de errores sutiles. Sabrás que ha funcionado cuando veas con claridad rasgos deformados, sombras extrañas o elementos sin lógica en la imagen.

💬 Ejemplo: ves una foto de una persona conocida en una situación inusual. Amplía sus manos y su cara. Si notas algo raro — dedos de más, rasgos deformados o una iluminación incoherente — investiga más antes de compartirla o de creerla.

Nota: los generadores de IA modernos han mejorado muchísimo, así que estos fallos son cada vez menos frecuentes. Cuando tengas dudas, verifica la imagen con una búsqueda inversa de imágenes u otros métodos.

Paso 2 — Identifica pistas de vídeos "deepfake"

Un **deepfake** es un vídeo en el que la cara o el cuerpo de una persona se ha alterado digitalmente, o incluso se ha intercambiado por completo, usando sistemas informáticos avanzados. Piensa en ello como una máscara digital sofisticada que se mueve en tiempo real e intenta imitar a la perfección a una persona real. Aunque detectarlos es cada vez más difícil a medida que mejora la tecnología, todavía hay algunas cosas que puedes vigilar.

Empieza reproduciendo el vídeo sospechoso y luego páusalo con frecuencia para observar bien algunos fotogramas concretos. Mientras reproduces el vídeo verás los fotogramas en movimiento y, al pausarlo, aparecerá una imagen fija que te permitirá analizar las zonas de la cara sin la distracción del movimiento constante. Sabrás que ha funcionado cuando detectes incoherencias en cómo la cara se une con el pelo, una iluminación rara o zonas ligeramente borrosas o deformadas.

💬 Ejemplo: recibes un mensaje de vídeo urgente del director de tu empresa pidiendo una transferencia inmediata de fondos. Pausa el vídeo y mira con atención la mandíbula, el cuello y la línea del pelo. Si notas algo raro — un contorno difuso, sombras inusuales o movimientos de labios que no encajan del todo con el audio — trátalo como sospechoso y verifícalo por otro canal antes de actuar.

Nota: los deepfakes de alta calidad son cada vez más difíciles de detectar a simple vista. Si un vídeo pide dinero o información sensible, verifica siempre por un método independiente y de confianza.

Paso 3 — Escucha los problemas del clonado de voz

El **clonado de voz** es una tecnología que toma una grabación corta de la voz de una persona real y luego usa un ordenador para hacerla decir cosas totalmente nuevas. Imagínalo como una herramienta avanzada de texto a voz capaz de imitar a la perfección a tu amigo, tu hijo o tu compañero de trabajo. El clonado de voz moderno puede ser muy convincente, así que la mejor protección no es fiarte solo de la voz.

Cuando recibas una llamada o un mensaje de audio sospechoso, escucha con atención la voz de la persona, fijándote bien en su tono y en su forma de hablar. Oirás las palabras que se dicen y podrás analizar si suenan naturales o no. Si estás en una llamada, intenta buscar un lugar tranquilo o usa auriculares si tienes a mano para reducir el ruido de fondo. Si algo te parece un

poco raro — un tono inusual, frases inesperadas o una petición que no encaja con la persona — tómalo como una señal para verificar por otros medios.

💬 Ejemplo: recibes una llamada urgente de un amigo que pide ayuda económica. La voz suena exactamente como la suya, pero la petición es inusual o el tono resulta extraño en ese contexto. Esto es una señal de alerta. Cuelga inmediatamente y llámalo a su número habitual y de confianza para verificar.

Paso 4 — Verifica los mensajes sospechosos usando un segundo canal

Los **mensajes engañosos** suelen aprovechar la IA para redactar correos electrónicos o mensajes de texto muy convincentes y personalizados, lo que los convierte en una forma potente de estafa automatizada. La regla de oro de la seguridad online es sencilla: nunca uses los datos de contacto que aparecen *dentro* de un mensaje sospechoso para verificar su autenticidad, porque seguramente te llevarán directo al estafador.

Tu acción debe ser alejarte a propósito del mensaje original e iniciar el contacto por otro método verificado de forma independiente. Verás el mensaje original en tu pantalla (correo, SMS, redes sociales), que puede contener solicitudes urgentes o datos "actualizados". Al usar un contacto independiente, evitas darle al estafador acceso directo. Si la solicitud llegó por correo electrónico, por ejemplo, no respondas a ese correo. En su lugar, busca un número de teléfono de confianza en una factura antigua o en la web oficial de la empresa. Sabrás que ha funcionado cuando la persona al otro lado confirme (o desmienta) la solicitud a través de un método de contacto que has iniciado tú y en el que confías, separado del que podría ser falso.

💬 Ejemplo: un proveedor envía un correo con "datos bancarios actualizados" para un pago y te urge a usar la nueva cuenta. No respondas a ese correo ni llames al número de teléfono que aparece en su firma. En su lugar, busca el número en tu contrato en papel original o en la web oficial del proveedor y llámalo directamente para confirmar el cambio.

Paso 5 — Busca textos genéricos o demasiado perfectos

El texto generado por IA muchas veces se identifica no por errores evidentes, sino por su *perfección* — o por la falta de matices humanos. Estos **textos escritos por IA** pueden ser demasiado formales, increíblemente verbosos, repetitivos en sus frases o sencillamente demasiado genéricos para parecer una comunicación humana genuina. Les faltan las peculiaridades, los matices emocionales y las pequeñas imperfecciones que hacen única la escritura humana.

Tu acción consiste en leer el mensaje o artículo con atención, fijándote bien en su estilo, su ritmo y la elección de palabras. Mientras lees, estarás evaluando el tono e identificando patrones

extraños. Esto se aplica tanto a un correo electrónico como a una publicación en redes sociales, un artículo online o incluso una reseña. Si el texto parece escrito por un robot que intenta sonar humano, es una buena pista. Sabrás que ha funcionado cuando la escritura se sienta anormalmente fluida, sin particularidades humanas ni coloquialismos, o use un lenguaje demasiado formal para una situación informal.

💬 Ejemplo: recibes un correo de un supuesto amigo, pero usa un lenguaje excesivamente formal como "Asimismo, necesito ayuda inmediata" en lugar de su estilo cercano habitual, y cada frase está perfectamente estructurada sin contracciones ni coloquialismos. Esa perfección tan poco natural sugiere intervención de una IA.

Errores comunes

- **Confiar en la imagen "oficial":** los estafadores pueden usar IA para copiar fácilmente logotipos de empresas, tipografías que parecen oficiales y lenguaje profesional. El error es dar por buena la legitimidad solo por la apariencia. **La solución:** céntrate siempre en la *acción* que te piden (como hacer clic en un enlace sospechoso o enviar dinero), no en lo profesional o auténtico que parezca el mensaje.
- **Dejarte llevar por el pánico y actuar demasiado rápido:** las estafas con IA están diseñadas para generar urgencia o miedo, y hacerte sentir presionado para actuar sin pensar. **La solución:** si un mensaje exige actuar de inmediato, respira hondo y haz una pausa. El tiempo siempre juega a tu favor cuando se trata de verificar hechos y protegerte.
- **No ver el contexto completo:** a veces se comparte un pequeño fragmento de una imagen, vídeo o audio fuera de su contexto original, lo que lo hace parecer sospechoso cuando en realidad es inofensivo. **La solución:** intenta siempre encontrar la fuente original o la historia completa si algo parece cuestionable. Busca en internet la imagen o el vídeo para ver dónde apareció por primera vez y cuál era su contexto original.

Qué significa esto para ti

- **En tu vida diaria:** puedes proteger a tu familia frente a las angustiosas llamadas de estafa entendiendo que una voz conocida al teléfono no siempre tiene por qué ser de quien esperas. Usar estas habilidades de observación te ayuda a verificar información importante antes de tomar decisiones precipitadas, y evita que tus seres queridos caigan en la explotación emocional y económica.
- **Para tu negocio o trabajo:** proteger los activos y la reputación de tu empresa resulta más fácil. Estarás mejor preparado para detectar correos de phishing, facturas fraudulentas o videollamadas deepfake que podrían provocar pérdidas económicas importantes o filtraciones de datos. Proteger a tu equipo y a tus clientes del engaño sofisticado impulsado por IA es clave para mantener la confianza y la seguridad.

- **Si estás empezando:** lo más importante es cultivar una buena dosis de escepticismo sobre el contenido que encuentras en internet, sobre todo si despierta una reacción emocional intensa o exige actuar con urgencia. Empieza por hacer una simple pausa y preguntarte: "¿Esto parece real?" antes de hacer clic, responder o compartir.

Prueba ahora

La próxima vez que estés mirando tus redes sociales o revisando tu correo, haz una pausa cuando veas una imagen, vídeo o mensaje que parezca demasiado bueno para ser verdad, o sorprendentemente fuera de lugar. Aplica una de las comprobaciones rápidas de esta guía: por ejemplo, amplía una foto o lee un mensaje en voz alta para detectar tonos robóticos.

Conclusión

Mantenerse informado y alerta es tu mejor defensa frente al panorama cambiante de las estafas y la desinformación impulsadas por IA. Al aprender a reconocer posibles señales de alerta en el contenido generado por IA, te capacitas a ti y a quienes te rodean para navegar por el mundo digital con más confianza y seguridad. Recuerda siempre cuestionar, verificar y confiar en tu instinto cuando algo no encaje.

Cómo detectar un deepfake o una voz clonada antes de compartirlo

5 pasos · Léela en la web

> **En resumen:** Fíjate en los ojos, la boca y los bordes de la cara en busca de pequeños defectos; escucha si la voz suena plana y robótica; comprueba quién lo envió y qué pide. Si algo te chirría, no lo reenvíes y verifica primero con la persona real por un canal de confianza.

Llega al grupo de la familia un vídeo corto de una figura pública diciendo algo escandaloso. Una nota de voz urgente de tu "jefe" pidiendo que compres tarjetas regalo. Nada de esto parece falso a primera vista, y ese es justo el truco. Un **deepfake** es un vídeo, imagen o audio creado con IA para imitar a una persona real que en realidad nunca dijo ni hizo eso, y la **clonación de voz** (reproducir la voz de alguien de forma convincente a partir de una muestra de audio corta) ya es fácil y barata. La buena noticia: con unos segundos de observación cuidadosa suele bastar para pillar las señales más comunes. Aun así, ninguna comprobación es infalible, por lo que tu última red de seguridad siempre es verificar con la persona real antes de compartir nada.

Antes de empezar

- El propio clip (guardado en el móvil u ordenador, o un enlace en un mensaje)
- Un navegador web (para la búsqueda inversa de imagen del paso 4)
- Unos 2-3 minutos por clip
- Sin cuenta, sin app y sin plan de pago

Paso 1 — Observa la cara en busca de pequeños defectos

Pausa el vídeo y estudia la cara durante unos 10 segundos. Céntrate en los **ojos, los bordes del rostro y la boca**. Señales habituales: ojos que no parpadean al ritmo correcto; gafas que parecen deslizarse o deformarse al girar la cabeza; dientes que se funden en un único bloque blanco al moverse la boca; un contorno de la cara que se ve borroso o "pintado" donde se junta con el pelo y el fondo; una luz en la cara que no encaja con el resto de la escena. Si no puedes pausar ni hacer zoom, baja la velocidad a 0,5x y vuelve a mirar. En el móvil, toca la pantalla para que aparezca la opción de velocidad de reproducción: normalmente es un pequeño botón "1x" cerca de la barra de tiempo. Sabrás que ha funcionado cuando localices al menos un detalle que te resulte extraño, o, igual de útil, cuando confirmes que la cara se ve natural.

Paso 2 — Escucha la voz en busca de un tono plano

Cierra los ojos un momento y simplemente escucha. La voz humana tiene pequeñas imperfecciones: una respiración rápida, un medio tartamudeo, una frase que baja de volumen

cuando el hablante duda. Las voces generadas por IA (TTS, abreviatura de text-to-speech: software que convierte texto escrito en audio hablado) suelen ser demasiado limpias, demasiado uniformes en volumen y curiosamente planas en emociones. Fíjate si hay frases sin ningún respiro, palabras apoyadas todas en el mismo tono, o un registro que no pega con la situación (alguien que se supone está asustado, pero suena calmado y sereno). En el móvil, usa auriculares para que el ruido de fondo no tape las pistas. Sabrás que ha funcionado cuando puedas distinguir si la voz suena humana o **sintetizada** (hecha por un ordenador, no por una persona real).

Paso 3 — Lee el mensaje que rodea al clip

Antes de reenviar nada, para y lee el contexto. ¿Quién lo ha enviado? ¿Es realmente esa persona, o solo un enlace reenviado por alguien de confianza? ¿Qué te pide el clip: donar dinero, pulsar un enlace, mandar un código, mantener algo en secreto? La mayoría de estafas se apoyan en la **urgencia, el secretismo o grandes afirmaciones emocionales**: "envíalo en la próxima hora", "no se lo cuentes a nadie", "esto es enorme". Son las mismas señales de alerta que en cualquier estafa, y son un buen motivo para ir más despacio. Si no puedes responder a "quién, qué y por qué" solo con la información del mensaje, eso ya es una señal. Sabrás que ha funcionado cuando tengas una respuesta clara a esas tres preguntas, o un motivo claro para desconfiar.

Paso 4 — Haz una comprobación rápida de la fuente

Haz una captura de la cara de la persona en el vídeo (o usa la opción "guardar imagen" de la plataforma). Abre un buscador inverso de imágenes (los más habituales son **Google Images**, **TinEye** o **Yandex Images**, no hace falta cuenta) y sube la foto. Busca apariciones anteriores de esa misma cara. Si se supone que es una figura pública y la búsqueda no devuelve nada, o si las fotos más antiguas no coinciden con la persona del clip, trátalo como un falso casi seguro. Si el clip supuestamente es de alguien que conoces, mándale un mensaje rápido por un canal que ya tenías antes de que apareciera el clip (no por el canal del que llegó el clip) y pregúntale si realmente lo envió. 💬 Ejemplo: "Oye, me ha llegado una nota de voz de alguien que dice ser tú preguntando por unas tarjetas regalo. ¿La has enviado tú? Solo quería comprobarlo." Sabrás que ha funcionado cuando puedas confirmar la fuente o decidir que no merece la pena arriesgarse.

Paso 5 — Decide qué hacer a continuación

Si alguna de las comprobaciones anteriores te ha generado dudas, lo más seguro es simple: **no lo compartas**. Reenviar un contenido falso, aunque sea con un "puede que sea fake", sigue difundiendo el original. Dile a quien te lo envió, con educación, que no has podido verificarlo y pregúntale de dónde lo sacó. Si el clip dice ser de una persona que conoces, avísala de que alguien puede estar usando su cara o su voz. Si el clip es un intento de estafa, repórtalo en la plataforma donde lo viste y, si se ha visto implicado dinero o el acceso a una cuenta, contacta con tu banco o

con la empresa suplantada directamente. Sabrás que ha funcionado cuando pares y pienses antes de pulsar "reenviar".

Errores comunes

- **Confiar en el encuadre de "comparte si estás de acuerdo".** Si un clip intenta que lo reenvíes antes de comprobarlo, eso ya es la estafa funcionando. Ve más despacio: el botón de compartir puede esperar 30 segundos.
- **Suponer que el nombre de un amigo significa que te lo envía un amigo.** Las cuentas se hackean y los contactos se pueden **suplantar** (un mensaje diseñado para parecer que viene de alguien que conoces). Verifica por otro canal antes de fiarte.
- **Apoyarte en una sola app de "detección de IA".** Ninguna app de consumo es lo bastante fiable como para ser tu única comprobación, y algunas herramientas de detección son en sí mismas una estafa. Tus ojos, tus oídos y una rápida comprobación de la fuente ganan a cualquier app sola.

Qué significa esto para ti

- **En tu día a día:** Trata como estafa, hasta que se demuestre lo contrario, cualquier vídeo o mensaje de voz inesperado que pida dinero, códigos o guardar secreto. Una comprobación de dos minutos protege tu cuenta bancaria y tu reputación.
- **En el trabajo:** Si un "compañero" te envía una solicitud rara por voz o vídeo, sobre todo si implica transferencias o credenciales, confírmala por un canal conocido, como una llamada de teléfono o una conversación en persona, antes de actuar.
- **Si estás empezando:** Guarda esta página y úsala como checklist la próxima vez que un clip te provoque una reacción fuerte, sea positiva o negativa. Las emociones intensas son justo lo que estos clips buscan provocar.

Pruébalo ahora

Elige un vídeo corto que hayas visto hace poco y que te haya provocado una reacción intensa. Dedica dos minutos a aplicarle los pasos 1, 2 y 4. Aunque resulte ser real, la práctica hará que la próxima comprobación sea más rápida y fácil.