Cloné mi voz con IA en 30 segundos. Mi madre habría pagado sin dudar.

El experimento que explica por qué tu familia necesita una palabra clave hoy, y por qué la ciberseguridad ya no puede depender de reconocer una voz

No hace falta ser ingeniero. No hace falta saber programar ni tener contactos oscuros en foros de la dark web. Cualquier adolescente con diez euros al mes y una tarde libre puede tener tu voz, exacta, diciendo lo que quiera, en el idioma que quiera, hoy mismo.

Lo sé porque lo hice con la mía.

El experimento: 30 segundos y 10 euros

Hace unas semanas hice un ejercicio que ya había retrasado demasiado. Quería comprobar, en primera persona, hasta dónde había llegado la democratización de la clonación de voz. No iba a documentar cómo se hace —eso sería irresponsable—, pero sí quería entender, y mostrar, qué significa que una tecnología que hace tres años estaba reservada a laboratorios especializados esté hoy al alcance de cualquier persona con una tarjeta.

Cogí un fragmento de una entrevista pública mía, de apenas 30 segundos. Nada raro: el tipo de audio que publicamos todos sin pensarlo cuando subimos un vídeo a LinkedIn o grabamos un podcast. Lo subí a una plataforma de síntesis de voz de uso general. En menos de un minuto tenía un modelo entrenado de mi voz.

Generé un audio corto. La frase era sencilla: *"Mamá, necesito que me hagas un Bizum urgente, no te puedo explicar ahora, luego te llamo."*

Lo escuché. Era yo. Mi acento, mis pausas, esa forma ligeramente atropellada de hablar cuando tengo prisa. No era "parecido". Era yo.

No llamé a mi madre. Ahí estaba la línea ética del experimento: la demostración educativa termina donde empezaría la prueba sobre una persona real que no ha dado su consentimiento. Pero me quedé con una certeza muy incómoda: si lo hubiera hecho, mi madre habría pagado sin dudar un segundo.

Por qué esto ya no es ciencia ficción

El caso corporativo más comentado en los últimos meses muestra hasta dónde ha llegado la sofisticación: un grupo vinculado a Corea del Norte utilizó una videollamada con un deepfake de la imagen y voz de un directivo conocido para convencer a la víctima de comprometer su seguridad durante una reunión de Zoom. No fue un correo sospechoso ni una web mal hecha. Fue un rostro familiar en una pantalla, diciendo lo que habría dicho el original.

Si esto ha llegado al nivel de actores estatales sofisticados en entornos corporativos blindados, la pregunta no es si llegará al fraude doméstico masivo. La pregunta es cuándo se normalizará.

Y esa normalización ya está aquí. Las llamadas de *"mamá, soy yo, he perdido el móvil, necesito dinero"* no son un mito urbano de 2018. Son una industria creciente que en 2026 ya no depende de imitadores hábiles ni de voces parecidas: depende de tres minutos de tu voz publicados en redes sociales.

Tu voz es materia pública (y no es culpa tuya)

Aquí hay algo que conviene interiorizar sin culpa: si tienes LinkedIn con vídeos, un podcast, entrevistas en YouTube, charlas grabadas, o incluso si has enviado notas de voz por WhatsApp que se han reenviado más allá de tu control, tu voz ya es material clonable. No es un fallo de seguridad tuyo. Es una consecuencia inevitable de vivir en un ecosistema digital donde la exposición es parte del juego profesional y social.

La respuesta instintiva —*"entonces dejo de publicar"*— no es realista ni efectiva. Primero, porque para la mayoría la presencia pública es parte del trabajo. Y segundo, porque basta un fragmento muy corto para entrenar un modelo razonable. No se puede borrar lo que ya está fuera.

Esto es importante entenderlo: la defensa no pasa por ocultar tu voz. Pasa por asumir que tu voz, como tu imagen, ya no puede ser en sí misma una prueba de identidad. Y si ya no lo es, hay que sustituirla por otra cosa.

El protocolo familiar: más útil que cualquier antivirus

La contramedida es antigua, analógica y desconcertantemente barata: una palabra de seguridad.

Una palabra o una frase corta, absurda, imposible de adivinar desde fuera, que solo conocen las personas que importan. No se comparte por chat. No se guarda en notas del móvil. No se menciona en voz alta en espacios públicos. Existe solo en la memoria de dos o tres personas.

Cuando alguien llame diciendo ser tu hijo en apuros, tu pareja pidiendo una transferencia urgente, o tu director general desde un aeropuerto con mala cobertura, hay una sola pregunta que necesitas hacer: *"Dime la palabra."* Si la dice, es real. Si no, cuelga y llama tú directamente al número que tienes guardado.

Esto parece trivial, casi cómico, hasta que entiendes lo que está neutralizando: toda una categoría de fraudes basados en suplantación de identidad por voz y vídeo, independientemente de lo sofisticada que sea la tecnología que los genere.

La palabra de seguridad es la razón por la que los protocolos sencillos suelen ganar a las defensas complejas. Un antivirus no te protege de que creas una voz. Una formación técnica avanzada tampoco te protege del vuelco emocional de oír a tu hijo llorando al teléfono. Pero una palabra acordada previamente, sí.

El cambio estratégico que 2026 nos obliga a hacer

Durante años hemos pensado la ciberseguridad como una cuestión de herramientas: mejores contraseñas, mejores antivirus, mejores firewalls. Esa lógica sigue siendo válida para la capa técnica, pero ya no es suficiente por sí sola.

La frontera se ha movido. Los ataques más peligrosos de 2026 no explotan vulnerabilidades en el software. Explotan vulnerabilidades en los protocolos humanos: la confianza por defecto, la respuesta emocional bajo presión, la suposición de que quien habla con la voz de alguien conocido es, efectivamente, esa persona.

Proteger a tu familia y a tu negocio hoy es, sobre todo, un ejercicio de diseño de protocolos. Qué se verifica antes de transferir dinero. Quién puede autorizar qué por qué canal. Qué palabra se pide antes de actuar sobre una llamada urgente. Qué se hace

cuando la urgencia choca con la verificación —y gana siempre la verificación—. No son medidas técnicas. Son cultura de verificación, aplicada en casa y en la oficina.

La buena noticia es que esta capa no requiere presupuesto, ni conocimientos informáticos, ni licencias anuales. Requiere una conversación de quince minutos con las personas que importan. Y probablemente es, a día de hoy, la inversión en ciberseguridad con mejor retorno que puedes hacer.

Tu próximo paso

Cierra este artículo y haz una cosa, solo una: decide una palabra de seguridad con tu madre, con tu pareja, con tus hijos, con tu socio. Una palabra rara, que no aparezca en tu biografía pública, que no tenga nada que ver con tu familia, tu trabajo ni tus aficiones más visibles. Dila en voz alta una vez. Acordad que nadie de fuera de ese círculo la escuchará jamás.

Esa palabra, que te costará treinta segundos elegir, va a valer más que cualquier consejo técnico que pueda darte en un artículo.

Si este texto te ha hecho pensar, compártelo con alguien que lo necesite: un padre, una pyme, un compañero de trabajo. Una familia con protocolo es una familia que el cibercrimen de 2026 no puede atacar con una voz clonada.

Isaac Ruiz Romero.