

DEEPFAKES: ¿CUÁLES SON SUS IMPLICACIONES SOCIALES Y PSICOLÓGICAS COMO NUEVAS FORMAS DE DESINFORMACIÓN?

Lic. Rodrigo Daniel Medrano Figueroa

Facultad de Estudios Superiores Iztacala, UNAM

ID

Resumen

Los deepfakes son videos, audios o imágenes que parecen reales, pero han sido creados con inteligencia artificial. Esta tecnología puede imitar rostros, voces y movimientos, haciendo que una persona parezca decir o hacer algo que nunca hizo. Aunque tienen usos creativos, su lado más oscuro genera problemas graves. Se utilizan para desinformar, manipular políticamente y difundir contenido dañino, como los deepnudes (Kopecký & Voráč, 2025), que violan la privacidad y causan daño emocional. La facilidad para crearlos y su rápida difusión en redes sociales complican la distinción entre lo real y lo falso. Esto afecta a la confianza pública y genera otras problemáticas. Además, la falta de regulaciones y de educación digital nos deja vulnerables. Por eso, el pensamiento crítico y la alfabetización digital son nuestra mejor defensa.

Palabras clave:

Deepfakes, Desinformación, Inteligencia artificial, Privacidad

Profesor de Asignatura Ordinario
Nivel "A" en la Facultad de Estudios Superiores Iztacala.
Correo electrónico:
rodirgo.medrano@iztacala.unam.mx



Recibido: 21 de abril de 2025

Aceptado: 26 de mayo de 2025

Cita recomendada:

Medrano, R. D. (2026). Deepfakes: ¿cuáles son sus implicaciones sociales y psicológicas como nuevas formas de desinformación? *Revista de Divulgación Crisis y Retos en la Familia y Pareja*, 8(1), 23-29.

<https://doi.org/10.22402/j.rdcfp.unam.8.1.2026.650.23-29>



ANÁLISIS

<https://doi.org/10.22402/j.rdcfp.unam.8.1.2026.650.23-29>



JUSTIFICACIÓN

¿Qué pasaría si no pudieras confiar en lo que ves o escuchas? Esta premisa se vuelve real en un mundo dominado por las redes sociales, donde la desinformación ha dejado de ser solo rumores o noticias falsas para convertirse en una amenaza mucho más sofisticada y peligrosa. Ejemplo de ello son los “deepfakes”, videos y audios creados con inteligencia artificial que manipulan o recrean rostros, voces y escenas para producir contenidos falsos con apariencia realista. Aunque esta tecnología está dirigida a aplicaciones creativas o educativas, en la práctica también se ha usado con fines maliciosos. De hecho, el problema se agrava porque estos contenidos pueden difundirse con gran rapidez en plataformas como TikTok, X (antes Twitter) o Facebook, y cada vez resulta más difícil distinguirlos de un material auténtico (Godulla et al., 2021). Por lo tanto, esta forma de manipulación no solo pone en duda nuestra percepción de la realidad, sino que también amenaza la confianza social, la privacidad e incluso el manejo de la información que consumimos a diario.

Por ello, resulta urgente comprender qué son los deepfakes, cómo se crean y cuáles son sus efectos más allá del ámbito tecnológico, ya que no se trata de un asunto exclusivo de especialistas, sino de un tema que nos concierne a todos, pues solo así podremos protegernos frente a los contenidos maliciosos que navegan por la red.

DESARROLLO

¿QUÉ SON LOS DEEPFAKES Y CÓMO FUNCIONAN?

Para entender este fenómeno, primero debemos saber de dónde viene su nombre. La palabra deepfake es la unión de dos conceptos clave: “deep learning” (aprendizaje profundo, una rama de la inteligencia artificial) y “fake” (falso).

En términos sencillos, un deepfake es un video, audio o fotografía que parece completamente real, pero que ha sido creado o manipulado artificialmente por un software de inteligencia artificial (IA) para engañar al público (Chawla, 2019). La IA los fabrica tan bien que resulta extremadamente difícil notar la manipulación.

Existen dos tipos principales de deepfakes que combinan la visión y el sonido:

- Sustitución de rostros (Deepfaces): La IA toma el rostro de una persona y lo “pega” sobre el cuerpo de otra en un video. El programa se asegura de que las expresiones faciales, el parpadeo y los movimientos de la boca coincidan perfectamente con el audio, creando una escena que nunca existió.
- Clonación de voz (Deepvoices): La IA aprende a imitar la voz de una persona con solo escuchar unos segundos de grabaciones. Una vez clonada, puede usar esa voz para decir cualquier cosa nueva,

generando audios y videos completos como si la persona original los hubiera grabado.

Al combinar estas técnicas, se pueden crear videos donde una figura pública (o cualquier persona) parece decir o hacer algo que jamás hizo. De hecho, estos montajes son tan realistas que la detección requiere herramientas especializadas.

LA TECNOLOGÍA DETRÁS DE LA ILUSIÓN

La magia detrás de los deepfakes se basa en un mecanismo de IA llamado Redes Generativas Antagónicas (GAN). Piensa en ellas como una competencia entre dos estudiantes de arte:

1. El Falsificador (Generador): Es el artista que crea un cuadro (el deepfake) lo más realista posible.
2. El Crítico (Discriminador): Es el evaluador que revisa el cuadro y dice: "Esto es falso" o "Esto es real".

Ambos trabajan juntos: el Falsificador intenta engañar al Crítico, y el Crítico se vuelve cada vez más astuto para detectar la falsedad. Esta constante competencia es lo que hace que los deepfakes sean cada día más perfectos y difíciles de diferenciar de la realidad (Ramos, 2024).

Hasta hace pocos años, estas manipulaciones solo las podían hacer expertos con acceso a hardware muy potente. Hoy, la tecnología se ha democratizado: cualquiera puede generar un deepfake de alta calidad en minutos usando una aplicación gratuita o una plataforma en línea (Lopez, 2022). Esta facilidad de acceso ha abierto la puerta a graves usos maliciosos que van más allá del simple engaño, afectando la estabilidad social y la salud mental de las personas.



RIESGOS SOCIALES Y CULTURALES

Uno de los problemas más preocupantes de los deepfakes es el deterioro de la confianza pública. Cuando la gente no puede distinguir entre lo verdadero y lo falso, se debilita nuestra capacidad para tomar decisiones informadas. Esta desconfianza no solo afecta el tipo de información que consumimos, sino también el ámbito político y las instituciones.

Por ejemplo, la difusión de un video falso de un candidato político haciendo declaraciones ofensivas justo antes de una elección puede provocar un cambio irreversible en la opinión pública, incluso si la falsificación se prueba después. Estos escenarios ya son una realidad en países como India, Estados Unidos y Brasil, donde los deepfakes se han usado para manipular audiencias y desprestigiar rivales políticos (Ramos, 2024).

Además, existe una variante especialmente grave: el ciberacoso por deepnudes (Kopecký & Voráč, 2025). Estos son vídeos o imágenes sexualizadas que reemplazan el rostro de una persona (casi siempre mujeres) en cuerpos desnudos. Esta forma de violencia digital afecta profundamente la integridad y la salud emocional de las víctimas, pues vulnera su privacidad y puede tener serias implicaciones personales, profesionales y sociales (Chawla, 2019).

Un estudio advierte que este contenido se ha convertido en una forma recurrente de acoso en línea, donde en muchos casos, la víctima no se entera del video manipulado hasta que ya ha sido difundido masivamente (El País, 2024).

IMPACTO EMOCIONAL Y PSICOLÓGICO

Más allá del daño a la reputación pública, los deepfakes tienen un impacto directo y profundo en la salud mental de las personas. La sola idea de que se pueda generar un contenido falso con el propio rostro o voz provoca ansiedad, desconfianza y una constante sensación de vulnerabilidad.

Esta crisis se vuelve mucho más intensa cuando los deepfakes circulan en entornos íntimos, donde el daño es personal y relacional:

- En contextos de pareja y familia: Un video o audio falsificado puede desencadenar una ruptura inmediata, conflictos severos o incluso violencia. El daño no es solo la mentira, sino la duda irreversible que instala: la víctima debe luchar por demostrar una inocencia que la mayoría de la gente da por sentada, y el solo hecho de que su pareja o familia sospechen causa un trauma emocional duradero.
- En el ámbito escolar y laboral: Los deepfakes se usan como una poderosa forma de bullying o humillación. El efecto es devastador porque el daño no radica únicamente en el contenido en sí, sino en las acciones y reacciones que genera: la exclusión social, el miedo al escrutinio público y el riesgo de que el montaje llegue a manos de profesores, jefes o reclutadores.

El problema es que el efecto devastador no se limita a la víctima. La poca o nula información que existe sobre esta tecnología agrava la situación.

Por ejemplo, según una encuesta realizada por Kaspersky en 2021 en México, el 72% de los encuestados no sabía cómo identificar un deepfake. Esto refleja una gran desventaja frente a una amenaza que se expande rápidamente.

Además, la falta de apoyo institucional refuerza el daño psicológico. La ausencia de recursos legales y psicológicos claros para las víctimas (ya que muchas personas no saben a dónde acudir ni cómo denunciar) las deja en un estado de desamparo. A esto se suma que los marcos jurídicos en muchos países todavía no se actualizan para tipificar claramente esta manipulación (Al-khazraji et al., 2023). Esta falta de protección oficial aumenta la sensación de soledad y vulnerabilidad de la víctima.

Del mismo modo, la falta de campañas informativas impide que el público desarrolle habilidades para identificar la manipulación digital. Esto no solo abre un panorama de vulnerabilidad donde cualquiera puede quedar expuesto públicamente, sino que también favorece la propagación de rumores y narrativas falsas que la gente acepta sin un ejercicio crítico. Esta situación es especialmente preocupante en entornos donde la información circula de manera acelerada y se consume sin un filtro, independientemente de que el contenido haya sido manipulado (Godulla, 2021).





¿CÓMO PODEMOS PROTEGERNOS?

A pesar del riesgo que conllevan, la protección frente a los deepfakes no es solo una cuestión tecnológica, sino también de estrategia emocional, social y legal. La defensa integral se basa en tres pilares:

1. Protección de la información y hábitos críticos

La primera línea de defensa es el escepticismo digital. Dado que el objetivo malicioso de estos contenidos es generar controversia o desinformación, es fundamental seguir estos pasos antes de reaccionar:

- Verificar la fuente: Siempre revisa quién publicó el material. Pregúntate: ¿Proviene de una cuenta confiable? ¿Ha sido confirmado por otras fuentes de noticias serias? Actuar con criterio es la mejor forma de no caer en el engaño.
- Buscar errores técnicos: Aunque son realistas, los deepfakes suelen fallar en detalles sutiles. Observa con detenimiento: ¿Los movimientos faciales son poco naturales? ¿Hay parpadeos irregulares o una iluminación inconsistente? ¿Hay desincronización entre los labios y la voz? (López, 2022).
- Detener la difusión impulsiva: Es crucial promover una cultura de verificación.

- Antes de compartir un video o audio en redes sociales o grupos privados, tómate un momento para cuestionar su autenticidad. No compartir es el primer acto de responsabilidad.

2. Afrontamiento emocional y apoyo relacional

La amenaza de los deepfakes genera vulnerabilidad, miedo y ansiedad. Por ello, proteger la integridad psicológica es tan vital como verificar la información.

Si eres víctima de un deepfake o conoces a alguien que lo es, la respuesta debe ser el apoyo y la acción legal, no la culpa. Es fundamental buscar redes de apoyo familiares y sociales para validar las emociones y enfrentar el problema colectivamente. No estás solo.

Además, es crucial saber que existen vías de acción formal:

- Asesoría legal y psicológica: Es importante acudir a autoridades o instituciones especializadas que brinden asesoría legal y acompañamiento psicológico. Muchas organizaciones no gubernamentales o instancias gubernamentales ya cuentan con protocolos para atender la violencia digital.
- Documentación: Guarda capturas de pantalla, enlaces y cualquier evidencia del contenido y su difusión. Esta documentación es indispensable para cualquier denuncia formal que se busque iniciar.

3. Exigencia de regulación y políticas

Finalmente, la protección requiere de un compromiso colectivo para mejorar el entorno digital. Es importante apoyar el desarrollo de iniciativas de regulación responsable que obliguen a las plataformas a actuar.

Así como existen políticas para combatir el discurso de odio, es importante generar medidas específicas que fuercen a las

plataformas digitales a implementar mecanismos eficaces para detectar y eliminar deepfakes maliciosos (Al-khazraji et al., 2023). Solo con protección legal, apoyo emocional y medidas de verificación se puede ofrecer una defensa integral frente a un problema que afecta tanto la esfera digital como la vida personal.

Conclusiones

Hemos visto a lo largo de este artículo que los deepfakes son mucho más que una simple "noticia falsa"; son un desafío que afecta directamente a nuestra confianza y a nuestra paz mental. La conclusión principal es clara: el riesgo no está solo en la tecnología, sino en la vida personal y las relaciones de la gente.

¿QUÉ NOS DEJA ESTE ANÁLISIS?

1. El daño es real y muy personal: El impacto más serio de los deepfakes no ocurre en la política, sino en la intimidad. Pueden romper parejas, generar acoso en el trabajo o la escuela, y causar un trauma psicológico que requiere ayuda. El lector debe saber que la víctima de un deepfake necesita apoyo emocional y legal, no ser juzgada.
2. La mejor defensa está en sus manos: No podemos esperar a que las leyes o las grandes plataformas resuelvan todo. La defensa más efectiva empieza por usted. Esto significa hacer dos cosas: desconfiar de lo que nos enoja o sorprende (verificar la fuente) y apoyar a quien sea víctima, sin compartir el contenido manipulado.
3. Hay que ser críticos, no ingenuos: Los deepfakes van a seguir mejorando. Por eso, el reto no es eliminarlos, sino aprender a vivir con ellos. La herramienta más poderosa que tenemos es el pensamiento crítico.

4. Al entender su impacto en nuestras vidas, convertimos la vulnerabilidad en una responsabilidad colectiva para protegernos y proteger a otros.

La era del deepfake nos obliga a ser ciudadanos digitales más astutos. Si no caemos en el engaño y si apoyamos a las víctimas, estaremos construyendo la mejor barrera contra este fenómeno.

REFERENCIAS

- Ajder, H. (2019). *The State of Deepfakes: Landscape, Threats, and Impact*. Deeptrace.
- Al-khazraji, S., Saleh, H., Khalid, A., & Mishkhal, I. (2023). Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications. *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, 23, 429-441. <http://www.epstem.net/en/pub/issue/79793/1371792>
- Chawla, R. (2019). Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*, 4(6), 4-8. <http://doi.org/10.22215/timreview/1282>
- El País. (2024). *El caso histórico de violencia digital con inteligencia artificial en México*. <https://elpais.com/mexico/2024-11-28/el-caso-historico-de-violencia-digital-con-inteligencia-artificial-en-mexico.html>
- Godulla, A., Hoffmann, C., & Seibert, D. (2021). Dealing with deepfakes, an interdisciplinary examination of the state of research and implications for communication studies. *SCM Studies in Communication and Media*, 10(1), 72-96. <http://dx.doi.org/10.5771/2192-4007-2021-1-72>
- Kaspersky. (2021). *La infodemia y su impacto en la vida digital*. kaspersky daily. <https://latam.kaspersky.com/blog/investigacion-la-infodemia-y-su-impacto-en-la-vida-digital>
- Kopecký, K., & Voráč, D. (2025). The phenomenon of deep nudes—a new threat to children and adults. *AI & Soc.* <https://doi.org/10.1007/s00146-025-02425-4>

López, J. (2022). Análisis y desarrollo de modelos de generación de voz mediante Deep Voice. Proyecto Fin de Carrera / Trabajo Fin de Grado, E.T.S. de Ingenieros Informáticos (UPM), Madrid, España. <https://oa.upm.es/71260/>

Ramos, F. (2024). Deepfake: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial. *Derecho global. Estudios sobre derecho y justicia*, 9(27), 359-387 <https://doi.org/10.32870/dgedj.v9i27.754>

