

Deepfakes





Introducción

Los avances de la inteligencia artificial permiten hoy crear o modificar imágenes, audios y videos con un nivel de realismo que dificulta distinguir lo auténtico de lo fabricado. Estos contenidos, tienen usos legítimos en la producción audiovisual y en la educación, pero también se emplean para suplantar identidades, hostigar y dañar la reputación de las personas. Es importante comprender que las instituciones educativas no son ajenas a esta problemática, ya que durante 2026 se conocieron en Argentina casos de deepfakes que involucraba a estudiantes. En el presente documento, explicaremos qué son los deepfakes, cómo identificarlos y verificarlos, qué riesgos implican para la identidad, la reputación y la convivencia escolar, y qué acciones de prevención y respuesta pueden adoptar las comunidades educativas.

¿Qué son los deepfakes?

La palabra deepfake combina dos términos ingleses: deep learning (aprendizaje profundo) y fake (falso). El aprendizaje profundo es una técnica de inteligencia artificial que permite a un sistema reconocer patrones complejos a partir de grandes cantidades de datos, como rostros, voces, movimientos, imágenes o textos (García-Ull, 2021; Kaspersky s.f.).

Un deepfake es un contenido audiovisual creado o modificado con inteligencia artificial para que una persona real parezca decir, hacer, mostrar o participar en algo que no ocurrió. Puede tratarse de una imagen, un audio, un video o una representación en tiempo real, como un avatar o una videollamada con el rostro alterado (Izaga Gallo, 2025; Kaspersky, s. f.).





Un sistema de IA puede analizar fotografías disponibles de una persona y generar una imagen nueva que la muestre en una situación inexistente. También puede estudiar grabaciones de su voz y producir un audio que imite su forma de hablar. En un video puede modificar expresiones faciales, sincronizar los labios con un audio distinto o construir una escena completa que parezca registrada por una cámara

La diferencia con la edición tradicional está en el grado de realismo y en que la herramienta genera elementos nuevos. No se limita a recortar una foto o aplicar un filtro: puede imitar rasgos, reconstruir escenas y producir sonidos que no existían.

Usos

Se usa en la educación como una herramienta para producción audiovisual, creación artística, simulaciones y educación. El problema aparece cuando se emplea para engañar, hostigar, estafar o dañar a otras personas. Entre los usos perjudiciales se encuentran:

- Suplantar la identidad de una persona.
 - Crear audios falsos atribuidos a docentes, autoridades o estudiantes.
 - Difundir información falsa o descontextualizada.
 - Manipular imágenes para burlarse, humillar o desacreditar.
 - Crear contenidos sin consentimiento.
 - Dañar la reputación de una persona o de una institución.
- 



Las comunidades educativas no están aisladas de lo que ocurre en internet. Estudiantes, docentes, directivos, familias y personal de apoyo usan redes sociales, grupos de mensajería, plataformas educativas y videollamadas. Por eso, un contenido manipulado puede afectar directamente la vida institucional.

Identificación de contenidos manipulados

La identificación se apoya en la combinación de varios indicios y, sobre todo, en la verificación de la fuente y del contexto. Ninguna señal alcanza por sí sola, sino que la detección resulta más confiable cuando se reúnen varias inconsistencias pequeñas y se combinan con el análisis del contexto (El Surtidor, 2023; Kaspersky, s. f.).

Algunos aspectos a tener en cuenta a la hora de identificar deepfakes son:

- Visuales: piel artificialmente lisa y sin imperfecciones, textura del cabello extraña, contornos borrosos, iluminación que no coincide con el resto de la escena. En cuanto a los ojos y boca se puede observar parpadeo escaso o excesivo, dientes poco definidos, expresiones exageradas o rígidas. Dedos deformes, objetos asimétricos. Imprecisiones en las formas, que se pueden presentar levemente distorsionadas (Gómez, 2026).
 - Movimientos poco naturales o entrecortados: solo se mueve la boca y el resto del cuerpo queda estático.
 - Desajuste entre la voz y el movimiento de los labios, distorsiones o ruido extraño
 - Falta de contexto: no hay fuente original identificable ni datos de origen. Apela a la
- 



urgencia, la indignación o el escándalo para que se reenvíe rápido.

Es importante resaltar que los generadores de deepfake mejoran constantemente y las señales visuales que hoy funcionan pueden dejar de servir mañana. Además, las herramientas automáticas de detección trabajan con probabilidades y también pueden fallar. Por eso, es importante formar el pensamiento crítico y la comprensión de cómo se produce y circula la información y cuáles son las responsabilidades.

Verificación responsable

Existen programas y plataformas que intentan detectar contenido generado por inteligencia artificial como Google Lens que se utiliza para comprobar si una imagen fue publicada en un contexto distinto al actual o Hive Detect que permiten arrastrar imágenes o videos cortos para evaluar la probabilidad de que hayan sido generados o alterados por inteligencia artificial. Estas herramientas trabajan con probabilidades y pueden equivocarse, de modo que un resultado automático no debe tomarse como prueba absoluta. Un contenido generado por IA puede no mostrar errores evidentes, y un material auténtico puede parecer sospechoso por su calidad. La verificación responsable combina varios pasos :

- Observar el contenido.
 - Revisar la fuente.
 - Analizar el contexto.
 - Comparar con otras fuentes.
 - Consultar a personas o instituciones confiables.
- 

- 
- Decidir no compartir si no se puede confirmar.

Riesgos para la identidad, la reputación y la convivencia

- **Identidad:** La identidad digital está formada por las imágenes, videos, publicaciones, comentarios y datos que circulan sobre una persona en internet. Cuantas más fotos, audios o videos públicos existan de una persona, más material hay disponible para imitar sus rasgos. Entonces, un deepfake afecta la identidad porque hace parecer que alguien estuvo en una situación, realizó una acción o expresó una opinión que nunca existió (Izaga Gallo, 2025). El contenido es falso, pero queda asociado a una persona real y condiciona cómo la perciben compañeros, docentes y familias. Que el material haya sido fabricado no reduce el daño: la exposición, la vergüenza, el miedo y la sensación de perder el control son reales. Tampoco corresponde responsabilizar a la persona afectada por tener fotos públicas o perfiles en redes ya que publicar una fotografía no autoriza a otras personas a copiarla, manipularla, ridiculizarla o usarla para suplantar una identidad. La responsabilidad recae en quien crea, difunde o utiliza el contenido de manera dañina.
 - **Reputación:** Un contenido falso puede instalar dudas sobre una persona incluso después de ser desmentido (García-Ull, 2021). El daño no proviene solo del contenido inicial, sino que lo potencian los comentarios, las reacciones y los reenvíos que lo acompañan contribuyen a producirlo.
- 

- **Convivencia escolar:** Es afectada por los conflictos entre estudiantes, el aislamiento de la persona afectada, las burlas y los insultos, el miedo a asistir a clases y las dificultades en los vínculos del curso (El Diario Paraná, 2026).

Casos en Argentina

Durante 2026, distintos medios argentinos informaron denuncias vinculadas con la creación y circulación de imágenes manipuladas con inteligencia artificial que afectaban a estudiantes de instituciones educativas (Página/12, 2026). Unos de los más relevantes son los casos del Colegio Nacional de Buenos Aires y Carlos Pellegrini donde estudiantes de ambos colegios fueron denunciados por usar y comercializar deepfakes de compañeras. Los acusados serían alumnos de 13 y 14 años. Todo empezó cuando alumnas hallaron un archivo compartido con sus nombres, fotos intervenidas. Además, se difundió una frase escrita en los bancos del aula que anunciaba que no dejarían de hacerlo.

Prevención

Prevenir no consiste en prohibir las redes sociales ni en pedir a los estudiantes que no publiquen fotos, sino en construir buenos hábitos. Como restringir el acceso a los perfiles de redes sociales para reducir las oportunidades de replicar su imagen o utilizar la función de "ver una sola vez" en WhatsApp para el envío de fotografías personales en chats grupales o con contactos poco frecuentes (Kaspersky, s. f.).



Los estudiantes necesitan entender que producir o compartir un contenido no es un acto aislado, sino que tiene consecuencias para sí mismos y para otras personas.

Además, es importante que desde las instituciones educativas informen a los alumnos a quién acudir dentro de la misma si reciben un contenido sospechoso o creen que alguien está siendo afectado. Que acompañen desde un Equipo de Apoyo y Orientación a quienes son afectados y concienticen a los estudiantes sobre las consecuencias y riesgos de crear, almacenar y difundir deepfakes.

Ante un deepfake:

- Se recomienda no reenviar, comentar, reaccionar de manera impulsiva y no pedir que otras personas manden el contenido. ***El primer objetivo es evitar la amplificación.***
 - Registrar datos mínimos para poder reportar: plataforma, usuario o cuenta, enlace, fecha y hora, y el contexto en que se recibió. No es necesario descargar, almacenar ni mostrar el material a más personas de las imprescindibles.
 - Intentar identificar el origen, es decir quién lo publicó primero, si hay una fuente institucional, si la fecha y el lugar son correctos y si otras fuentes confiables confirman la información.
 - Informar la situación a adultos o autoridades institucionales.
- 



Conclusión

Los deepfakes representan un desafío creciente para las instituciones educativas, porque pueden producirse y difundirse con facilidad y generan consecuencias reales en las personas afectadas, aunque el contenido sea falso. Dado que ninguna señal visual ni herramienta de detección es infalible, la verificación de la fuente y del contexto, junto con el pensamiento crítico, constituye la principal estrategia de protección.

Por ello, la prevención debe apoyarse en hábitos de cuidado de la información personal, en la formación sobre las consecuencias de crear y difundir este tipo de material y en espacios claros de acompañamiento para quienes resultan afectados. La responsabilidad recae siempre en quien crea o difunde el contenido, y nunca en quien lo padece. Pasar del compromiso a la acción implica que toda la comunidad educativa actúe de manera informada, coordinada y empática.





Referencias

Gómez J. (2 ene 2026) *Los deepfakes inundan tus redes: Por qué los videos de IA son peligrosos y cómo detectarlos.* DW español.

<https://www.youtube.com/watch?v=S2IU150Fb60>

Gómez J. (9 mar 2020) *¿Qué son los deepfakes? Los trucos para detectar las falsificaciones de videos.* DW español.

<https://www.youtube.com/watch?v=ACVglWBYhQA>

El Argentino Diario. (2026, 13 de julio). *Alumnas del Nacional y Pellegrini, víctimas de deepfakes pornográficos vendidos entre compañeros.*

<https://elargentinodiario.com.ar/politica/educacion-2/13/07/2026/alumnas-del-nacional-y-pellegrini-victimas-de-deepfakes-pornograficos-vendidos-entre-companeros/>

El Diario Paraná. (2026, 11 de julio). *Desnudos con IA en colegios de élite: vejación real, imagen falsa.* <https://www.eldiario.com.ar/2026/07/11/desnudos-con-ia-en-colegios-de-elite-vejacion-real-imagen-falsa/>

<https://www.eldiario.com.ar/2026/07/11/desnudos-con-ia-en-colegios-de-elite-vejacion-real-imagen-falsa/>





El Surtidor. (2023, 22 de mayo). *Guía para identificar imágenes y videos generados con inteligencia artificial*. <https://elsurti.com/chequeo/2023/05/22/guia-para-identificar-imagenes-y-videos-generados-con-inteligencia-artificial/>

García-Ull, F. J. (2021). *DeepFakes: El próximo desafío en la detección de noticias falsas*. *Anàlisi*, (64), 103–120. <https://doi.org/10.5565/rev/analisi.3378>

Infobae. (2026a, 9 de julio). *Denuncian a alumnos de secundario por generar fotos de sus compañeras desnudas con IA y venderlas*. <https://www.infobae.com/sociedad/2026/07/09/denuncian-a-alumnos-de-secundario-por-generar-fotos-de-sus-companeras-desnudas-con-ia-y-venderlas/>

Infobae. (2026b, 8 de agosto). *«Nos pueden delatar, pero no vamos a parar...» (packs de \$1.000, causa por explotación sexual infantil)*. <https://www.infobae.com/sociedad/policiales/2026/08/08/nos-pueden-delatar-pero-no-vamos-a-parar-desnudaban-a-sus-companeras-con-ia-y-vendian-sus-fotos-por-packs-de-mil-pesos/>

Izaga Gallo, A. (2025). *La metamorfosis de la verdad: Deepfakes y el desafío de la autenticidad en la sociedad digital*. *Derecom. Revista Internacional de Derecho*





de la Comunicación y de las Nuevas Tecnologías, (35), 1-18.

<https://dx.doi.org/10.5209/dere.102344>

Kaspersky. (s. f.) *Protegete de los deepfakes: ¿Qué son y cómo detectarlos?*

<https://latam.kaspersky.com/resource-center/threats/protect-yourself-from-deep-fake>

Página/12. (2026, 9 de julio). *Varios alumnos de secundarios de la UBA fueron acusados de crear deepfakes con IA de sus compañeras y vender las fotos.*

<https://www.pagina12.com.ar/2026/07/09/varios-alumnos-de-secundarios-de-la-uba-fueron-acusados-de-crear-deepfakes-con-ia-de-sus-companeras-y-vender-las-fotos/>

