REVISTA DE **DERECHO PRIVADO:** DEBATES Y TENDENCIAS
Núm. 02 (2025) · ISSN 3087-2332· doi 10.5354/3087-2332.2025.80910
Recibido: 01/10/2025 · Aprobado: 20/11/2025 · Publicado: 31/12/2025

Deepfakes íntimos no consentidos: desafíos del ordenamiento jurídico chileno y el rol de la responsabilidad civil ante la IA

Non-Consensual Intimate Deepfakes: Challenges for the Chilean Legal System and the Role of Civil Liability in the Face of AI

Vanesa Jabbaz Rosenbaum
Universidad de Chile, Chile

RESUMEN: Este artículo examina el fenómeno del *deepfake* íntimo no consentido en Chile, con especial atención a la responsabilidad civil. Se revisan experiencias comparadas en la Unión Europea, Estados Unidos, China y Australia, con el fin de identificar tendencias regulatorias relevantes a nivel internacional. Asimismo, se incorporan referencias al marco constitucional, penal y a la legislación especial para contextualizar el entorno normativo general. Finalmente, se analiza si la legislación civil vigente resulta suficiente para reparar los daños derivados de los *deepfakes* en general y de los íntimos no consentidos en particular. El artículo concluye que no existe un vacío normativo, correspondiendo aplicar principios generales e internacionales, y formula propuestas de *lege ferenda* orientadas a fortalecer el sistema de responsabilidad civil.

PALABRAS CLAVE: *deepfake*, responsabilidad civil, derecho privado, violencia digital, inteligencia artificial.

ABSTRACT: This article examines the phenomenon of non-consensual intimate deepfakes in Chile, with a particular emphasis on civil liability. It reviews comparative experiences in the European Union, United States, China, and Australia in order to identify regulatory trends of international relevance. References to constitutional, criminal and special legislation are included to contextualize the general legal framework. Finally, the article assesses whether the existing civil law legislation is sufficient to remedy the harms arising from deepfakes in general, and from non-consensual intimate deepfakes in particular. The article concludes that there is no normative

gap, since general and international principles remain applicable, while also putting forward de *lege ferenda* proposals to strengthen the system of civil liability.

KEYWORDS: deepfake, civil liability, private law, digital violence, artificial intelligence.

Introducción

La irrupción de las tecnologías de inteligencia artificial ("IA") ha generado oportunidades inéditas, pero también riesgos que desafían de manera directa al derecho. Entre estas tecnologías, los *deepfakes* —contenidos audiovisuales manipulados mediante algoritmos de aprendizaje profundo y redes neuronales, capaces de producir falsificaciones altamente verosímiles— constituyen uno de los fenómenos más disruptivos de los últimos años (Chesney y Citron, 2019: 1759).

Si bien esta técnica puede destinarse a fines benignos, como el entretenimiento o la educación, su utilización con propósitos lesivos ha puesto en jaque a los sistemas jurídicos. Entre los diversos usos problemáticos, el *deepfake* íntimo no consentido[1] se erige como uno de los más graves por su frecuencia, alcance y efectos. La combinación de difusión masiva en redes sociales y la facilidad de acceso a herramientas de creación hace que este fenómeno atente contra bienes jurídicos esenciales como la dignidad, la honra y la intimidad de las personas.

La respuesta normativa al fenómeno de los *deepfakes* ha sido diversa. En el derecho comparado, tanto la Unión Europea como países como Estados Unidos, China y Australia han implementado marcos regulatorios y leyes que buscan prevenir y enfrentar el impacto de estas prácticas, incluidas aquellas de carácter sexual. En Chile, un caso ocurrido en un conocido colegio de Santiago (en adelante, el "Colegio

1 En la literatura académica y en los medios de comunicación se utiliza con frecuencia la expresión "pornografía *deepfake*" o "*deepfake* pornográfico" para referirse a la manipulación de imágenes o videos sexuales mediante inteligencia artificial sin consentimiento de la persona afectada. En este trabajo se opta por la expresión "*deepfake* íntimo no consentido", por considerarse más precisa e inclusiva: abarca no solo el material de carácter pornográfico en sentido estricto, sino también desnudos parciales u otras representaciones íntimas, y pone el énfasis en la ausencia de consentimiento como elemento central del problema jurídico.

de Santiago"), en que estudiantes difundieron imágenes pornográficas falsas de compañeras menores de edad, puso en evidencia la urgencia de esta discusión. Si bien el ordenamiento contempla mecanismos desde el derecho constitucional, penal y leyes especiales, tales respuestas resultan todavía limitadas para enfrentar los desafíos específicos que plantean los *deepfakes*.

En este contexto, desde la perspectiva del derecho privado —y en particular la responsabilidad civil— resulta importante examinar cómo las categorías clásicas de voluntariedad, culpa, causalidad y daño se proyectan frente a este fenómeno. El objetivo de este trabajo no es sólo evaluar si dichas categorías ofrecen respuestas adecuadas, sino también explorar si es necesario avanzar hacia nuevas formas de atribución de responsabilidad y reparación del daño.

En consecuencia, este artículo se propone analizar el fenómeno de los *deepfakes*, con especial atención a su utilización en contextos sexuales no consentidos. El desarrollo se organiza en tres etapas: primero, se revisan experiencias del derecho comparado, con el objeto de identificar las soluciones adoptadas en otros sistemas jurídicos; luego, se evalúa si el ordenamiento jurídico nacional ofrece herramientas suficientes para responder a este fenómeno; y, finalmente, se analiza en detalle el rol de la responsabilidad civil, evaluando tanto la suficiencia de sus categorías tradicionales como la necesidad de repensarlas. Sobre esta base, se formulan propuestas de *lege ferenda* destinadas a fortalecer la capacidad del sistema jurídico chileno para ofrecer respuestas eficaces y coherentes frente a los daños derivados de esta práctica.

El fenómeno del *deepfake* y sus principales problemas

El término *deepfake* proviene de la combinación de *deep learning* y *fake*, y alude a la creación de contenido multimedia manipulado utilizando la interacción de redes neuronales avanzadas y algoritmos de aprendizaje profundo (Fletcher, 2018:458). Esta tecnología permite sintetizar rostros, modificar expresiones o incluso suplantar identidades completas, generando la apariencia de que una persona ha dicho o realizado acciones que nunca ocurrieron en la realidad (Tolosana y otros, 2020:

132). La IA y el aprendizaje profundo son esenciales en este proceso, pues permiten una precisión y realismo sorprendentes que dificultan la identificación de su falsedad.[2]

Los *deepfakes* suelen materializarse en videos, imágenes o audios que imitan la apariencia y la voz de una persona, al punto de engañar tanto a individuos como a algoritmos. El procedimiento habitual implica usar imágenes de dos individuos en un algoritmo de aprendizaje profundo. Este algoritmo, utilizando tecnología de mapeo facial e IA, se entrena para reemplazar el rostro de una persona en una fotografía o video por el de otra (Westerlund, 2019: 40).

Aunque existe un mercado creciente de aplicaciones de consumo que emplean tecnología *deepfake* para el entretenimiento —como *FaceSwap*[3]—, era solo cuestión de tiempo que esta tecnología se utilizara con fines perjudiciales. No se trata de una simple edición de imágenes y videos, sino derechamente de falsificaciones altamente realistas producidas con IA. El usuario solo ingresa un *prompt* (instrucción o comando) para que el sistema de IA genere el contenido deseado, sin necesidad de involucrarse en los detalles técnicos de la creación.[4]

Además, los *deepfakes* evolucionan continuamente para burlar la percepción de la audiencia o de otros algoritmos con inteligencia artificial que se encargan de detectar esta tecnología. Para ello se utiliza el aprendizaje profundo y las redes generativas antagónicas (GAN), término acuñado por Goodfellow y otros (2014), donde dos redes neuronales compiten entre sí: el generador que produce datos artificiales similares a los datos reales, mientras que el discriminador analiza las salidas creadas por el generador e identifica si son artificiales o reales. El objetivo es que el generador cree contenidos lo suficientemente realistas para

---

2 Uno de los ejemplos más emblemáticos es el del vídeo en el que aparece Barack Obama insultando a Donald Trump. Véase "Barack Obama Appears to Insult Donald Trump" (video de YouTube). Disponible en https://www.youtube.com/watch?v=cQ54GDm1eL0, consultado el 3 de septiembre de 2025.

3 Tecnología que permite intercambiar los rostros de dos personas en imágenes o videos.

4 C. Solís y M. Baricco, "Inteligencia artificial y derecho de daños: el fenómeno *Deepfake* y los vehículos autónomos", Ciclo de videoconferencias del Colegio Público de la Abogacía de la Capital Federal, Buenos Aires, 12 de diciembre de 2023 (video de YouTube). Disponible en https://www.youtube.com/watch?v=Ktl42fRajt4, consultado el 10 de agosto de 2025.

engañar al discriminador, hasta volver las falsificaciones prácticamente indistinguibles de las auténticas (Goodfellow y otros, 2014: 1).

En este contexto, los *deepfakes* han comenzado a plantear problemas jurídicos y morales de gran envergadura.[5] Uno de los más visibles es su utilización en la difusión de desinformación y noticias falsas, especialmente en contextos políticos y electorales. El conocido caso del político indio Manoj Tiwari en 2020, en el que circuló un video manipulado en el que aparentemente hablaba en un dialecto local con fines de propaganda electoral,[6] demuestra cómo esta tecnología puede ser instrumentalizada para alterar la percepción de la opinión pública.

Otro problema radica en la afectación de los derechos de la personalidad, como la privacidad, la honra y la propia imagen. En 2020, el rapero Jay-Z demandó a la plataforma de YouTube *Vocal Synthesis* por un *deepfake* de audio en el que se le hacía recitar pasajes de Shakespeare sin su consentimiento, evidenciando cómo la explotación de la voz y la imagen de una persona puede vulnerar gravemente su identidad y sus derechos.[7]

Los *deepfakes* plantean también serias dificultades desde la perspectiva de la propiedad intelectual e industrial, al permitir la creación de obras derivadas no autorizadas que imitan o explotan el trabajo de artistas y creadores sin su permiso. Asimismo, en ámbitos financieros y de seguridad, la tecnología ha sido utilizada para suplantar identidades con fines de fraude, aumentando el riesgo de engaños y estafas en entornos digitales.

En suma, el fenómeno del *deepfake* presenta un doble rostro: por un lado, herramientas con potencial innovador en educación, entretenimiento y producción audiovisual; por otro, un uso lesivo que amenaza

---

5 Sensity. "The State of Deepfakes 2024." Informe disponible en https://sensity.ai/reports/, consultado el 1 de septiembre de 2025.

6 "BJP Shared Deepfake Video On WhatsApp During Delhi Campaign", NDTV. Disponible en https://www.ndtv.com/india-news/in-bjps-deepfake-video-shared-on-whatsapp-manoj-tiwari-speaks-in-2-languages-2182923, consultado el 1 de septiembre de 2025.

7 Jay-Z demanda a un canal de YouTube por suplantar su voz para simular lecturas de Shakespeare a través de Inteligencia Artificial." *Legal Army Blog*. Disponible en https://www.legalarmy.net/blog/jay-z-demanda-a-un-canal-de-youtube-por-suplantar-su-voz-para-simular-lecturas-de-shakespeare-a-traves-de-inteligencia-artificial, consultado el 10 de agosto de 2025.

bienes jurídicos fundamentales como la dignidad, la intimidad y la honra.

En este panorama de riesgos y desafíos, resulta evidente que no todos los usos problemáticos de los *deepfakes* tienen la misma magnitud ni generan el mismo nivel de daño. Entre ellos, el *deepfake* íntimo no consentido constituye, sin duda, el fenómeno más extendido y dañino. Ejemplo de ello fue el caso de la actriz Gal Gadot, cuando en la plataforma Reddit comenzó a circular un supuesto video pornográfico creado mediante *deepfake*. Dicho video, que mostraba imágenes de la actriz en escenas explícitas, fue producido con una aplicación que utiliza IA capaz de combinar imágenes y videos en movimiento de manera altamente realista.[8]

El *deepfake* íntimo no consentido no se limita a figuras públicas. También afecta a mujeres comunes, incluidas menores de edad, como se ha evidenciado en casos escolares[9] en España, México, Estados Unidosy, recientemente, en Chile, en el caso del Colegio de Santiago donde estudiantes generaron y difundieron imágenes falsas de sus compañeras menores de edad mediante tecnologías de *deepfake*.[10] A diferencia de otros usos —como la manipulación política, la suplantación de identidad o la desinformación—, el *deepfake* íntimo no consentido incide de manera directa en la dignidad y en la autonomía sexual de las personas, vulnera gravemente su privacidad y produce daños morales de muy difícil reparación. A ello se suma la facilidad con que estas imágenes o videos pueden difundirse a través de redes sociales y plataformas digitales, lo que amplifica exponencialmente sus efectos y genera un círculo de revictimización constante.

---

8 "Deepfake: los riesgos tras la técnica digital de moda y cómo detectarlos", Universidad Autónoma de Chile, 9 de abril de 2021. Disponible en https://www.uautonoma.cl/noticias/deepfake-los-riesgos-tras-la-tecnica-digital-de-moda-y-como-detectarlos/, consultado el 2 de septiembre de 2025.

9 "Deepfake porn, la Inteligencia Artificial da nueva cara al ciberacoso escolar", CIECEM – Simposio. Disponible en https://2024.ciecem.org/ponencia/deepfake-porn-la-inteligencia-artificial-da-nueva-cara-al-ciberacoso-escolar/, consultado el 10 de septiembre de 2025.

10 "Alumnos del Colegio Saint George crearon y viralizaron fotos de sus compañeras desnudas usando IA: Fiscalía investiga denuncia", *The Clinic*, 24 de mayo de 2024. Disponible en https://www.theclinic.cl/2024/05/24/alumnos-del-saint-george-viralizan-fotos-de-sus-companeras-desnudas-hechas-con-ia/, consultado el 10 de septiembre de 2025.

De este modo, el análisis jurídico del fenómeno *deepfake* no puede prescindir de un examen específico sobre su uso sexual. Este constituye un eje central del debate normativo actual en el derecho comparado y plantea interrogantes urgentes para el ordenamiento chileno: ¿están nuestras normas legales en condiciones de dar una respuesta adecuada? ¿o resulta necesario un rediseño normativo que reconozca las particularidades de esta práctica?

Con estas interrogantes en mente, el próximo capítulo abordará la regulación del *deepfake* a nivel comparado, explorando la forma en que distintos sistemas jurídicos han intentado enfrentarlo, y los desafíos que tales experiencias proyectan sobre el debate chileno.

Deepfake en el derecho comparado

El fenómeno del *deepfake* ha suscitado preocupación normativa a nivel global por la variedad de riesgos que plantea: desde la desinformación y la manipulación política hasta la suplantación de identidad, el fraude y la vulneración de derechos de autor. Sin embargo, entre sus múltiples manifestaciones, el *deepfake* íntimo no consentido constituye el aspecto más extendido y lesivo, dado que más del 98 % de los contenidos *deepfake* disponibles en línea corresponden a material sexual y que el 99 % de las víctimas son mujeres.[11] Se trata de un problema que involucra bienes jurídicos fundamentales —dignidad, autonomía sexual, privacidad, identidad— y que profundiza la vulnerabilidad estructural frente a la violencia de género (Simó, 2023). En su estudio, Simó retoma la tesis de Wagner y Blewer (2019: 33) sobre cómo estos cosifican el cuerpo femenino como objeto visual, ignorando el consentimiento.

A continuación, se revisan las principales respuestas regulatorias que distintos sistemas jurídicos han desarrollado frente a este fenómeno en general, con especial atención a las medidas adoptadas para abordar su dimensión íntima no consentida.

---

11 Home Security Heroes. "State of Deepfakes 2023." Informe disponible en https://www.homesecurityheroes.com/state-of-deepfakes/#appendix, consultado el 10 de septiembre de 2025.

Unión Europea

La Unión Europea se ha consolidado como referente mundial en la regulación de la IA y, dentro de ella, de los *deepfakes*.[12]

En 2018, la Comisión Europea advirtió el problema de la desinformación con su declaración *Tracking online disinformation: a European Approach*,[13] que fijó principios para prevenir la manipulación de la opinión pública por parte de los medios de comunicación. Esta declaración respondió a la creciente preocupación sobre la desinformación y las *fake news*, incluyendo los *deepfakes*, en las plataformas digitales. Entre sus lineamientos se encuentran: aumentar la transparencia en la producción y difusión de información; fomentar la alfabetización mediática para que los ciudadanos puedan identificar y resistir la desinformación; desarrollar tecnologías de detección y verificación; y fortalecer la colaboración entre gobiernos, plataformas y medios de comunicación para enfrentar el fenómeno de manera coordinada.

Posteriormente, la Unión Europea aprobó el Reglamento General de Protección de Datos (*General Data Protection Regulation,* GDPR),[14] que estableció un marco robusto para la protección de los datos personales, aplicable también a tecnologías avanzadas como la síntesis profunda (*deep synthesis technology*), esencial en la creación de *deepfakes*.

---

12 En la Unión Europea, el término "*deepfake*" suele traducirse como "ultrasuplantación" o "falsificación profunda", conforme a la terminología empleada por la Comisión Europea y parte de la doctrina especializada.

13 Comisión Europea. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Tackling Online Disinformation – A European Approach.* COM(2018) 236 final, Bruselas, 26 de abril de 2018. Disponible en https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236, consultado el 20 de julio de 2023.

14 Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). Disponible en https://eur-lex.europa.eu/eli/reg/2016/679/oj, consultado el 20 de septiembre de 2025.

Un paso decisivo en la regulación de contenidos en línea fue la Ley de Servicios Digitales (*Digital Services Act,* DSA) de 2022,[15] que moderniza la Directiva sobre comercio electrónico en materias de contenidos ilegales, publicidad transparente y desinformación. Este instrumento refuerza la responsabilidad de las plataformas digitales en la gestión de contenidos falsos y manipulados, entre ellos los *deepfakes*.

La piedra angular, sin embargo, es el Reglamento (UE) 2024/1689 de Inteligencia Artificial (*AI Act*), aprobado en junio de 2024 y en vigor desde agosto del mismo año. A diferencia de una Directiva—que requiere ser incorporada a la legislación nacional por cada Estado miembro—, el *AI Act* es un Reglamento de aplicación directa, tal como explica la Comisión Europea, porque "los reglamentos [...] adquieren automáticamente carácter vinculante en toda la UE a partir de su fecha de entrada en vigor".[16]

En efecto, se trata de la primera normativa integral a nivel mundial que regula el diseño, desarrollo y uso de sistemas de IA. Su objetivo es

> mejorar el funcionamiento del mercado interior mediante el establecimiento de un marco jurídico uniforme [...] a fin de promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta de los Derechos Fundamentales de la Unión Europea.[17]

---

15 Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo, de 19 de octubre de 2022, relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Ley de Servicios Digitales). Disponible en https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065, consultado el 20 de septiembre de 2025.

16 Comisión Europea. Implementación del Derecho de la Unión: reglamentos, directivas y su aplicación en los Estados miembros. Disponible en https://commission.europa.eu/law/application-eu-law/implementing-eu-law, consultado el 22 de noviembre de 2025.

17 Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de marzo de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), art. 1.1, Disponible en https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32024R1689, consultado el 16 de agosto de 2025.

En sus borradores iniciales (2021-2023), el *AI Act* distinguía cuatro categorías de riesgo (inaceptable, alto, limitado y mínimo o nulo).[18] En el texto definitivo esta clasificación se simplificó, manteniéndose explícitamente sistemas de riesgo inaceptable (art. 5), que amenazan gravemente los derechos fundamentales y están prohibidos, como aquellos que manipulan el comportamiento humano o explotan vulnerabilidades de grupos específicos. Aquí podrían incluirse los *deepfakes* íntimos no consentidos que afecten a mujeres, niños, niñas y adolescente. El *AI Act* también regula los sistemas de alto riesgo (art. 6 y Anexo III), que impactan negativamente la seguridad o los derechos fundamentales, sometiéndolos a evaluación y certificación estricta; por ejemplo, los *deepfakes* usados para influir en procesos electorales o democráticos.

Para otros supuestos, como los *deepfakes* con fines de entretenimiento o publicidad, el *AI Act* impone obligaciones de transparencia (art. 50), exigiendo que se identifique de manera clara y visible que se trata de contenidos generados o manipulados artificialmente, y contempla un deber reforzado cuando el material representa a una persona identificable. El objetivo es evitar el engaño y resguardar derechos como la identidad, la dignidad y la intimidad.

Como explica Estella (2025: 271-272), estas obligaciones recaen en dos figuras diferenciadas por el Reglamento: el proveedor y el responsable del despliegue.[19] Esta distinción es relevante, porque en el caso específico de los *deepfakes*, el art. 50.4 asigna la obligación de informar

---

18 Comisión Europea. Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021) 206 final, 2021/0106(COD)), 21 de abril de 2021. Disponible en https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52021PC0206, consultado el 16 de agosto de 2025.

19 Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de marzo de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), arts. 3.3 y 3.4, que definen respectivamente al proveedor como "una persona física o jurídica, autoridad pública, órgano u organismo que desarrolle un sistema de IA o modelo de IA de uso general y lo introduzca en el mercado (…) previo pago o gratuitamente", y al responsable del despliegue como la "persona física o jurídica, o autoridad pública, órgano u organismo que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional".

y "hacer público" su carácter artificial exclusivamente al responsable del despliegue, y no al proveedor.

A juicio de Estella (2025: 278), esta obligación se concreta en un deber de etiquetado visible que permita a los usuarios reconocer que están ante una "ultrasuplantación". Además, advierte que las obligaciones del art. 50 operan de manera acumulativa, de modo que un *deepfake* puede estar simultáneamente sujeto a marcado técnico, comunicación y etiquetado (Estella, 2025: 278-280). Sin embargo, el Reglamento no determina cómo debe implementarse dicha etiqueta, ni establece mecanismos eficaces para asegurar que los usuarios efectivamente la apliquen, lo que constituye —como señala el propio autor— una importante deficiencia regulatoria (Estella, 2025: 280-282).

El *IA Act* complementa instrumentos previos como la Resolución del Parlamento Europeo N. 53 de 2017[20] sobre de tecnologías emergentes, pero con la novedad de ofrecer un marco específico y detallado para la IA. Asimismo, se alinea con las recomendaciones internacionales de la Organización para la Cooperación y el Desarrollo Económico (OCDE)[21] y por la Organización de las Naciones Unidas para la Educación (UNESCO) sobre IA,[22] orientadas a garantizar que el desarrollo y uso de la IA respeten los derechos humanos y promuevan principios de transparencia, diversidad, credibilidad e inclusividad en el uso y desarrollo de tecnologías avanzadas.

Finalmente, pese a su carácter de marco uniforme para la UE, la eficacia dependerá de su articulación con legislaciones nacionales de los Estados miembros, que deberán dictar normas complementarias. En consecuencia, este Reglamento se proyecta como una hoja de ruta normativa para prevenir los riesgos inherentes a la IA, incluidos los *deepfakes* —en particular los de carácter íntimo no consentido—, al tiempo

20 Parlamento Europeo. Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)). Disponible en https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52017IP0051&from=EN, consultado el 16 de septiembre de 2025.
21 OCDE. Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449), 21 de mayo de 2019. Disponible en https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449, consultado el 16 de septiembre de 2025.
22 UNESCO. Recomendación sobre la ética de la inteligencia artificial, 23 de noviembre de 2021. Disponible en https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa, consultado el 16 de septiembre de 2025.

que fomenta la innovación y un desarrollo tecnológico centrado en las personas.

Estados Unidos

En Estados Unidos, la regulación de los *deepfakes* ha seguido un camino mixto, combinando respuestas estatales y federales.

En el ámbito estatal, California prohíbe la distribución de *deepfakes* destinados a dañar la reputación de una persona o a influir en procesos electorales en los 60 días previos a una elección;[23] Virginia, por su parte, penaliza la difusión no consentida de imágenes y videos sexualmente explícitos, incluyendo *deepfakes*.[24] Otros Estados también han avanzado en legislaciones específicas: Tennessee promulgó en 2024 la *Ensuring Likeness, Voice, and Image Security Act* (ELVIS Act),[25] orientada a proteger a los artistas frente al uso no autorizado de su voz, imagen y semejanza mediante inteligencia artificial; y en 2025, Washington aprobó la *House Bill 1205*, que criminaliza de manera amplia los *deepfakes* maliciosos, estableciendo tanto responsabilidad civil como sanciones penales para su uso ilegítimo.[26]

A nivel federal, destacan algunos hitos. El *Malicious Deepfake Prohibition Act of 2018* fue la primera norma en definir formalmente el concepto de *deepfake*.[27] Posteriormente, el *Deepfakes Accountability Act* (2019) buscó imponer deberes de identificación y transparencia a creadores y distribuidores de este tipo de contenidos.[28] Ese mismo año, el *Deepfake Report Act (*2019*)* ordenó al Departamento de Seguridad Nacional elaborar

---

23 California Civil Code, Section 1708.86, relativa a la prohibición de *deepfakes* íntimos no consentidos.
24 Virginia Code, Section 18.2-386.2, sobre creación y difusión de imágenes íntimas falsas.
25 Tennessee General Assembly, *Ensuring Likeness, Voice, and Image Security Act* (ELVIS Act), Public Chapter No. 734
26 State of Washington Legislature, *House Bill 1205-S* (Session Laws 2025–2026).
27 *Malicious Deepfake Prohibition Act of 2018, S.3805, 115th Congress (2018*), Sec. 2: "an audiovisual record created or altered in a manner that the record would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of an individual". Disponible en https://www.congress.gov/115/bills/s3805/BILLS-115s3805is.pdf, consultado el 16 de agosto de 2025.
28 *Deepfakes Accountability Act*, H.R.5586, 116th Congress (2019).

informes periódicos sobre los riesgos asociados a estas tecnologías.[29] Más recientemente, en mayo de 2025, el presidente Donald Trump promulgó la *Take It Down Act (TIDA),*[30] que penaliza la difusión de pornografía de venganza —ya sea real o generada por IA— y obliga a las plataformas digitales a retirar el contenido en un plazo máximo de 48 horas tras la solicitud de la víctima, reforzando así la protección frente al *deepfake* íntimo no consentido.

China

En China también se han implementado normativas específicas para enfrentar la creación y difusión de *deepfakes*, enmarcadas en una política más amplia de control del ecosistema digital. Un hito fue la promulgación de las "*Provisions on the Governance of the Online Information Content Ecosystem*", las cuales imponen a las plataformas de contenido en línea a gestionar y eliminar contenidos falsos o dañinos, incluyendo los *deepfakes*. Estas disposiciones obligan, además, a etiquetar de manera visible los contenidos manipulados, mantener registros de los mismos y reportarlos a las autoridades competentes, reforzando así el control estatal sobre la circulación de información digital.[31]

En noviembre de 2022, la *Cyberspace Administration of China (CAC)* emitió las *"Provisions on the Administration of Deep Synthesis Internet Information Services"*, que entraron en vigor el 10 de enero de 2023. Esta normativa estableció un marco regulatorio integral para los servicios de síntesis profunda, imponiendo deberes específicos a proveedores, desarrolladores y usuarios. Entre sus principales exigencias se cuentan la verificación de identidad real cuando se alteren atributos biométricos

---

29 *Deepfake Report Act of 2019*, S.2065, 116th Congress (2019).
30 U.S. Congress. *S.146 – Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act* (*Take It Down Act*), 119th Congress (2025–2026), aprobado por ambas cámaras y promulgado el 19 de mayo de 2025. Disponible en https://www.congress.gov/bill/119th-congress/senate-bill/146, consultado el 16 de agosto de 2025.
31 *Provisions on the Governance of the Online Information Content Ecosystem*, Order of the Cyberspace Administration of China (No. 5), 1 de marzo de 2021, traducción y comentario de Bolin Zhang y Joan Barat. Disponible en https://wilmap.stanford.edu/entries/provisions-governance-online-information-content-ecosystem, consultado el 16 de septiembre de 2025.

como rostro o voz, la obligación de etiquetar de forma clara los contenidos generados mediante síntesis profunda, y la implementación de medidas técnicas para evitar que dichos contenidos engañen al público.

Asimismo, los proveedores deben mantener registros, eliminar información ilegal o falsa y reportar incidentes a las autoridades competentes, todo ello con el fin de reforzar la transparencia y proteger derechos fundamentales como la privacidad y la identidad de las personas.[32] Estas reglas buscan no solo combatir la desinformación, sino también limitar la creación y difusión de *deepfakes* sexuales no consentidos, imponiendo sanciones administrativas y, en casos graves, responsabilidades penales.

La experiencia china se caracteriza, por tanto, por un enfoque marcadamente preventivo y centralizado, que combina la protección de los derechos de las personas frente a falsificaciones digitales con el control político del espacio virtual.

Australia

En Australia, la atención se ha centrado particularmente en el *deepfake* íntimo no consentido, considerado una forma de violencia digital con fuerte impacto en la dignidad, la autonomía y la privacidad de las víctimas. Este fenómeno se enmarca dentro de la problemática más amplia del *image-based abuse*, que en los últimos años ha adquirido creciente reconocimiento como una forma de violencia de género facilitada por tecnología.

En el plano federal, un avance decisivo fue la aprobación del *Criminal Code Amendment (Deepfake Sexual Material) Act 2024 (Cth)*, promulgado

---

32 Cyberspace Administration of China (CAC). *Provisions on the Administration of Deep Synthesis Internet Information Services*, adoptadas el 25 de noviembre de 2022 y en vigor desde el 10 de enero de 2023. Traducción y análisis en DigiChina, "Translation: Provisions on the Administration of Deep Synthesis Internet Information Services" (2022). Disponible en https://digichina.stanford.edu/work/translation-provisions-on-the-administration-of-deep-synthesis-internet-information-services/, consultado el 16 de septiembre de 2025. Véase también Latham & Watkins. *China's New AI Regulations: China's regulations aim to address risks related to artificial intelligence and introduce compliance obligations on entities engaged in AI-related business*, 16 de agosto de 2023. Disponible en https://www.lw.com/admin/upload/SiteAttachments/Chinas-New-AI-Regulations.pdf, consultado el 16 de septiembre de 2025.

el 2 de septiembre de 2024.[33] Esta reforma introdujo el nuevo artículo 474.17A al *Criminal Code Act 1995*, criminalizando expresamente la creación y distribución no consentida de material sexualmente explícito generado o alterado mediante IA. La norma sanciona tanto la producción como la difusión de esos contenidos, estableciendo una pena máxima de seis años de prisión para la conducta básica y hasta siete años en sus formas agravadas, cuando, por ejemplo, el autor haya manipulado directamente el material o cuente con sanciones previas en materia de seguridad en línea. Lo relevante de este tipo penal es que contempla explícitamente los *deepfakes*, estableciendo que "es irrelevante si el material transmitido está en forma sin alteración o si ha sido creado, o alterado de cualquier manera, mediante tecnología".[34] En otras palabras, declara irrelevante la naturaleza del material —real o fabricado digitalmente—, reconociendo que el daño a la víctima existe incluso cuando la representación íntima es producto de síntesis profunda. Con ello, el legislador federal buscó cerrar vacíos legales y dar una señal clara de que el *deepfake* sexual no consentido constituye una forma de abuso digital grave y autónomamente sancionada.

En paralelo, diversos estados australianos han avanzado en la tipificación de conductas asociadas. En Victoria, el *Crimes Act 1958 (Vic)* —versión autorizada a 10 de febrero de 2025[35]— contempla expresamente delitos relacionados con imágenes íntimas no consentidas incluyendo aquellas alteradas digitalmente. Por ejemplo, bajo la división titulada *"Producing, distributing or threatening to distribute intimate images"* (delimitada como *Subdivision (8FAAB)*) incorpora los siguientes artículos clave: 53R (*Producing intimate image*), 53S (*Distributing intimate image*) y 53T (*Threat to distribute intimate image*), todos los cuales regulan conductas relativas a imágenes íntimas sin consentimiento, sean

33 Australia. *Criminal Code Amendment (Deepfake Sexual Material) Bill 2024 (Bill R7205)*, introducido el 5 de junio de 2024 y sancionado el 2 de septiembre de 2024. Disponible en https://parlinfo.aph.gov.au/parlInfo/download/legislation/bills/r7205_aspassed/toc_pdf/24071b01.pdf, consultado el 21 de septiembre de 2025.
34 Australia. *Criminal Code Amendment (Deepfake Sexual Material) Bill 2024 (Bill R7205)*, Sec. 474.17A(2).
35 Australia, *Crimes Act 1958 (Vic)*, No. 6231 of 1958, versión autorizada con enmiendas incorporadas al 10 de febrero de 2025. Disponible en https://content.legislation.vic.gov.au/sites/default/files/2025-02/58-6231aa307-authorised.pdf, consultado el 16 de septiembre de 2025.

reales, manipuladas digitalmente, o simuladas en cuanto aparentan contenido íntimo. Este conjunto normativo permite que material generado o alterado digitalmente —lo que cabría identificar como *deepfake* íntimo no consentido— encaje dentro del tipo penal vigente en Victoria, aun cuando el término "*deepfake*" no se menciona literalmente en los artículos.

En Queensland, el *Criminal Code Act 1899* también contempla disposiciones aplicables a los *deepfakes* de carácter sexual. En particular, la sección 223 tipifica como delito la distribución de imágenes íntimas sin consentimiento, sancionándola con hasta tres años de prisión.[36] Lo relevante es que, conforme a la definición de *intimate image* prevista en la sección 207A, el concepto incluye no solo representaciones reales, sino también aquellas que hayan sido alteradas o directamente creadas para aparentar desnudez o actos sexuales.[37] De este modo, la normativa de Queensland permite subsumir los *deepfakes* íntimos no consentidos dentro de las figuras penales vigentes, incluso en ausencia de una referencia explícita al término *deepfake*.

Finalmente, cabe destacar que en Nueva Gales del Sur (NSW) se presentó en agosto de 2025 el *Crimes Amendment (Deepfake Sexual Material) Bill 2025 (NSW)*, actualmente en trámite parlamentario.[38] El proyecto busca modificar el *Crimes Act 1900 (NSW)* para extender los delitos relativos a imágenes íntimas no consentidas a aquellas creadas o alteradas digitalmente, incluyendo material audiovisual y audio sintético de carácter sexual. El texto fue introducido en la *Legislative Assembly* el 8 de agosto de 2025 y se encuentra en su segundo debate ("*2R Debate*"), pendiente de aprobación.

El caso australiano evidencia una doble vía de protección: por un lado, el impulso de una legislación federal que otorgue coherencia y

---

36 Australia. *Criminal Code Act 1899 (Qld)*, Sec. 223: "Distributing intimate image", disponible en https://www.legislation.qld.gov.au/view/whole/html/inforce/current/act-1899-009#sec.207A, consultado el 16 de septiembre de 2025.
37 Australia. *Criminal Code Act 1899 (Qld)*, Sec. 207A: "Definitions for this chapter", disponible en https://www.legislation.qld.gov.au/view/whole/html/inforce/current/act-1899-009#sec.207A, consultado el 16 de septiembre de 2025.
38 Parliament of New South Wales. *Crimes Amendment (Deepfake Sexual Material) Bill 2025 (NSW)*, disponible en https://www.parliament.nsw.gov.au/bills/Pages/bill-details.aspx?pk=18782, consultado el 16 de septiembre de 2025.

uniformidad al tratamiento del fenómeno; y por otro, la existencia de marcos estatales que ya ofrecen respuestas concretas frente al *deepfake* sexual no consentido, particularmente en el ámbito penal.

El caso chileno: respuestas normativas frente al *deepfake*

A diferencia de lo ocurrido en otras jurisdicciones —como Australia, Estados Unidos o la Unión Europea—, Chile no cuenta aún con legislación específica sobre *deepfakes*. El vacío normativo se hace patente en casos recientes, como el ocurrido en el Colegio de Santiago, que han revelado la urgencia de revisar si nuestro ordenamiento jurídico dispone de herramientas suficientes para responder a este fenómeno o si, por el contrario, resulta necesario avanzar en reformas legislativas o la dictación de normas especiales.[39]

En este contexto, surge una pregunta central que orienta el análisis de esta sección: ¿requiere Chile una legislación específica sobre deepfakes, o basta con normas de alcance general y tecnológicamente neutrales que protejan frente a la creación y difusión de contenido íntimo no consentido, cualquiera sea la tecnología utilizada**?** Si bien este trabajo no pretende resolver de manera exhaustiva esta interrogante, sí busca delinear los elementos que permiten evaluar la suficiencia del marco vigente y detectar sus principales límites. Con esta premisa, a continuación se examinan las respuestas actuales del ordenamiento chileno desde distintas áreas del derecho.

---

39 Un desafío particularmente complejo para las víctimas de *deepfakes* íntimos no consentidos es el problema de la jurisdicción internacional. Con frecuencia, el contenido manipulado es alojado, procesado o distribuido desde servidores ubicados fuera de Chile, lo que reduce la eficacia de los mecanismos nacionales de tutela y dificulta la obtención de órdenes de retiro o bloqueo. En la práctica, muchas afectadas recurren a mecanismos alternativos de bajada de contenido basados en derechos de propiedad intelectual —principalmente la *Digital Millennium Copyright Act* (DMCA) estadounidense— que, pese a no estar diseñados para la protección de derechos de la personalidad, han demostrado ser más expeditos y efectivos para lograr la desindexación o eliminación de material en plataformas globales. Esta asimetría evidencia tanto la insuficiencia de las herramientas nacionales como la necesidad de avanzar hacia protocolos transfronterizos de cooperación, capaces de enfrentar fenómenos digitales inherentemente globales.

Derecho constitucional: la vía del recurso de protección

En el ordenamiento chileno, la herramienta de tutela inmediata frente a vulneraciones de derechos fundamentales es el recurso de protección. En principio, esta acción podría invocarse para amparar la integridad psíquica, la igualdad ante la ley y los derechos de imagen,[40] privacidad y honra, garantizados en el artículo 19 N° 1, 2 y 4 de la Constitución Política de la República. Sin embargo, su eficacia práctica frente a los *deepfakes* resulta sumamente limitada, pues la viralización rápida e irreversible del material impide una tutela eficaz que logre hacer frente a las consecuencias de esta práctica.

El caso del Colegio de Santiago es ilustrativo. Allí, un grupo de apoderados presentó un recurso de protección ante la Corte de Apelaciones de Santiago denunciando que la entidad educacional no activó los protocolos ni aplicó las sanciones que correspondían según su reglamento interno a los alumnos responsables La sentencia acogió el recurso y reconoció la afectación de derechos fundamentales, exigiendo que el colegio ajustara sus medidas disciplinarias a su reglamento interno y, en definitiva, expulsara a los alumnos responsables. La Corte sostuvo que los hechos vulneraban el artículo 19 N°1 de la Constitución, "en la medida que mantener a los alumnos agresores, dentro del mismo recinto educacional al que asisten las afectadas, causa un daño psíquico a las víctimas, hijas de los recurrentes", y también el artículo 19 N°2, pues

> discrimina a las adolescentes, respecto de otras personas que, en idénticas condiciones, obtienen la aplicación de la medida disciplinaria vinculada a la gravedad de la falta cometida, que

40 Corte Suprema, 10 de noviembre de 2015, rol N° 9973-2015, considerando quinto: "[...] en lo tocante al resguardo constitucional del derecho a la propia imagen, a que precisamente tiende la acción propuesta en autos, es cierto que el artículo 20 de la Carta Fundamental no lo enumera determinadamente entre las garantías susceptibles de ampararse por ese arbitrio cautelar; empero, tanto la doctrina, como la jurisprudencia coinciden en que su protección deviene procedente y encuadra en el artículo 19 N° 4 de la Constitución, por encontrarse implícitamente comprendida en el atributo de privacidad de la persona, que esa norma se encarga de tutelar". En el mismo sentido, véanse Corte Suprema, sentencias roles N° 238.387-2023, N° 17.160-2023 y N° 152.338-2022.

> determina, además, que víctimas y agresores, no deban seguir conviviendo en el mismo espacio de formación escolar.[41]

Es decir, en este caso las medidas se orientaron a tutelar los derechos fundamentales que se veían comprometidos por la falta de respuesta institucional, pero no existió ninguna acción judicial encaminada a prevenir o detener la creación y difusión del *deepfake* sexual en sí mismo. De este modo, la protección se aplicó *ex post,* sin hacerse cargo de la amenaza, perturbación o privación directa que genera este tipo de actos en el ejercicio de derechos fundamentales.

Por otra parte, cabe preguntarse si las medidas que pueden ordenarse en virtud de un recurso de protección —como ordenar el retiro de las imágenes, prohibir futuras publicaciones o declarar su falsedad— permiten realmente una reparación del daño causado. La experiencia demuestra que la naturaleza viral e irreversible de la difusión digital hace prácticamente imposible contener el daño reputacional, psicológico y social de las víctimas. En consecuencia, esta acción cumple más bien una función cautelar, al decretar medidas inmediatas de protección, pero no asegura una reparación integral, lo que pone de relieve la necesidad de mecanismos regulatorios más eficaces frente a los desafíos del entorno digital.

A este límite se suma un debate conceptual: ¿los *deepfakes* vulneran efectivamente el derecho a la propia imagen? La Corte Suprema ha entendido este derecho como "una proyección física de la persona, que le imprime un sello de singularidad distintiva" y que permite a cada individuo controlar cuándo y cómo se captan y reproducen sus rasgos fisonómicos.[42] Sin embargo, los *deepfakes* no reproducen necesariamente estos rasgos de forma fidedigna, sino que los manipulan o crean imágenes completamente nuevas. Surge, entonces, la duda de si estas representaciones adulteradas pueden subsumirse dentro de la noción tradicional de imagen.

En definitiva, el recurso de protección en Chile ofrece un marco de tutela parcial frente al *deepfake*: es un mecanismo expedito para obtener medidas inmediatas de protección frente a la vulneración

41 Corte de Apelaciones de Santiago, 9 de mayo de 2024, rol N° 13.557-2024.
42 Corte Suprema, 9 de junio de 2009, rol N° 2506-2009.

de garantías fundamentales como consecuencia de esta tecnología, pero no asegura una reparación integral de los daños causados. Esta debe buscarse mediante el ejercicio de otras acciones administrativas o judiciales —civiles o penales— que permitan no solo sancionar a los responsables, sino también otorgar indemnizaciones y medidas de reparación adecuadas.

Derecho penal: insuficiencias y límites

Si bien no es objeto de este trabajo abordar de manera exhaustiva las implicancias penales asociadas a la producción y difusión de *deepfakes*, resulta pertinente formular algunas consideraciones generales que permiten evidenciar las insuficiencias del marco penal vigente.

En materia penal, la producción y difusión de *deepfakes* íntimos no consentidos no se encuentra tipificada en el ordenamiento chileno. Ante esta ausencia, podría pensarse en reconducir los hechos hacia tipos ya vigentes, tales como el delito de injurias, pornografía infantil o el de trato degradante a menores.[43] Sin embargo, todos estos intentos enfrentan serias dificultades derivadas del principio de tipicidad, a lo que se suma la complejidad de atribuir responsabilidad penal en casos donde el contenido es generado por sistemas de inteligencia artificial que no son sujetos de imputación penal.

Por otra parte, en mayo de 2019 se adoptó la Ley N° 21.153,[44] que modificó el artículo 161-C del Código Penal, incorporando el delito de captación y difusión no consentida de imágenes íntimas en espacios públicos o de libre acceso público. Conforme a su texto, esta norma sanciona a quien capte, grabe, filme o fotografíe, por cualquier medio, imágenes de los genitales u otra parte íntima del cuerpo de una persona, con fines de significación sexual y sin su consentimiento, siempre que ello ocurra en lugares públicos o de libre acceso público. Del mismo modo, penaliza a quien difunda dichas imágenes o registros, imponiendo

43 Código Penal chileno, artículos 416 y siguientes (injurias); artículos 367 ter y siguientes, en especial el artículo 367 quinquies (pornografía infantil); y artículo 403 ter en relación con el artículo 403 bis (trato degradante a menores).
44 Ley N° 21.153 de 2019, que modifica el Código Penal para tipificar el delito de acoso sexual en espacios públicos.

una pena superior cuando la persona es simultáneamente quien obtiene y divulga el material.

Para efectos de lo que interesa en este trabajo, el alcance del artículo 161-C aparece limitado, en la medida que el tipo penal exige una captación real y no consentida de imágenes íntimas en un espacio público o de libre acceso público. Bajo esta configuración, y atendida la estructura típica, parecería excluirse la posibilidad de aplicar esta norma a supuestos en los que el contenido no proviene de una captación efectiva, sino de una fabricación digital mediante técnicas de inteligencia artificial. En mi opinión, ello dificulta extender el artículo 161-C a los *deepfakes*, puesto que en estos casos no existe una imagen auténtica captada en un entorno físico, sino un producto sintético que reproduce o simula la corporalidad de la víctima, lo que desborda los límites del precepto en su formulación actual.

Por último, debe considerarse el artículo 161-D del Código Penal, incorporado por la Ley N.º 21.675,[45] en su artículo 56 Nº 2, que añadió esta disposición a continuación del artículo 161-C. El precepto sanciona a quien, sin autorización expresa, exhiba un registro de imágenes o sonidos que represente una acción sexual que involucre a otra persona o imágenes íntimas de connotación sexual, cualquiera sea su forma de obtención. Asimismo, establece una penalidad más severa cuando el registro es enviado, difundido o publicado. A diferencia del artículo 161-C, esta norma no exige que las imágenes provengan de una captación real ni que hayan sido obtenidas en un espacio público o de libre acceso público. Sin embargo, subsiste la cuestión interpretativa relativa a si un *deepfake* puede ser calificado como un "registro" en los términos del tipo penal, considerando que se trata de un producto sintético, generado mediante técnicas de inteligencia artificial, y no de una captación efectiva de la corporalidad de la víctima. Esta ambigüedad limita la posibilidad de aplicar el artículo 161-D a los *deepfakes* íntimos no consentidos y confirma que el marco penal vigente no fue diseñado para abordar la problemática específica de estos contenidos.

45 Ley Nº 21.675 de 2024, que estatuye medidas para prevenir, sancionar y erradicar la violencia contra las mujeres en razón de su género.

En conclusión, desde la perspectiva penal comparada, diversas jurisdicciones han optado por crear tipos autónomos para abordar la manipulación y creación sintética de imágenes íntimas, justamente para superar los vacíos derivados de la estricta exigencia de tipicidad. En contraste, en Chile subsiste una orfandad normativa, en la que solo caben interpretaciones forzadas de figuras penales preexistentes, lo que genera incertidumbre y desprotección para las víctimas.

Leyes especiales sobre violencia de género y violencia digital

En el plano legislativo, Chile ha avanzado en la construcción de un marco normativo que, aunque aún insuficiente, abre la puerta a reconocer fenómenos como los *deepfakes* íntimos no consentidos bajo la categoría de violencia digital y de género.

Por una parte, tras siete años de tramitación, en mayo de 2024 se promulgó la Ley N° 21.675, conocida como Ley Integral contra la Violencia hacia las Mujeres. Su objeto es prevenir, sancionar y erradicar la violencia contra toda mujer en razón de su género, incorporando medidas de prevención, protección, atención, reparación y de acceso a la justicia para ellas.[46–47]

Si bien el legislador perdió la oportunidad de incluir expresamente la violencia digital dentro del artículo 6°, que describe formas de violencia de género, esta disposición no es taxativa, lo que permite extender su alcance a otras formas de violencia no contempladas de manera explícita.[48] Por otra parte, el *deepfake* íntimo no consentido puede ser comprendido dentro de la noción de violencia simbólica, definida como

46 Ley N° 21.675 de 2024, que estatuye medidas para prevenir, sancionar y erradicar la violencia contra las mujeres en razón de su género.
47 Entre sus efectos, la Ley Integral contempla medidas de prevención, protección judicial y reparación, fortalece las medidas cautelares, crea unidades especializadas y refuerza la coordinación interinstitucional. En materia sancionatoria, modifica diversas normas del Código Penal y otras leyes, incorporando sanciones que van desde multas hasta penas de presidio, según la gravedad del hecho y la forma de violencia ejercida.
48 Ley N° 21.675, artículo 6: "Formas de violencia de género. La violencia en contra de las mujeres en razón de su género incluye, entre otras, las siguientes: (...)".

> toda comunicación o difusión de mensajes, textos, sonidos o imágenes en cualquier medio de comunicación o plataforma, cuyo objeto sea naturalizar estereotipos que afecten su dignidad, justificar o naturalizar relaciones de subordinación, desigualdad o discriminación contra la mujer que le produzcan afectación o menoscabo.[49]

Por otra parte, se encuentra en tramitación el Proyecto de Ley, Boletín N° 13.928-07, ingresado en diciembre de 2020 y actualmente en segundo trámite constitucional en el Senado.[50] Esta iniciativa busca fortalecer el marco penal frente a la violencia digital,[51] tipificando conductas que afectan la dignidad, intimidad y seguridad de las personas, con especial atención al entorno digital y a su dimensión de género.[52] A la fecha, el avance del proyecto ha sido tratado con "sentido de urgencia" por la Comisión de la Mujer y Equidad de Género, conformándose además una mesa técnica y convocándose audiencias con diversos

---

49 Ley Nº 21.675, artículo 6, numeral 5.

50 Este proyecto tiene como objetivo "prevenir, sancionar y erradicar la violencia digital y otorgar protección a las víctimas de la misma" mediante la tipificación de conductas graves en entornos digitales, como la difusión no consentida de contenido íntimo y el hostigamiento digital. Propone sanciones efectivas —incluyendo presidio menor en su grado mínimo (61 a 540 días) y multas de hasta 20 UTM— para la exhibición, el envío o la publicación de material íntimo sin consentimiento y para el hostigamiento persistente que perturbe gravemente la vida privada o la integridad psíquica de la víctima. Asimismo, incorpora la figura del *doxing* —difusión de datos personales que permitan ubicar físicamente a la víctima— y prevé agravantes en casos de víctimas menores de 14 años, relaciones de parentesco o pareja, anonimato o suplantación, y cuando la conducta persiga fines extorsivos (Boletín N.º 13.928-07, Proyecto de ley sobre violencia digital, Biblioteca del Congreso Nacional de Chile. Disponible en: https://www.camara.cl/legislacion/ProyectosDeLey/tramitacion.aspx?prmID=14490, consultado el 29 de septiembre de 2025).

51 Conforme al artículo 161-E del proyecto, se entenderá por violencia digital: "toda conducta realizada a través de tecnologías de la información y las comunicaciones, tales como medios, plataformas o dispositivos tecnológicos, que atente contra la integridad, la intimidad, la libertad o la vida privada, y que cause daño o sufrimiento psicológico, físico, económico, sexual, o a la identidad o expresión de género, tanto en el ámbito privado como en el público".

52 Biblioteca del Congreso Nacional de Chile, Nuevos delitos de violencia digital. Observaciones al Proyecto de Ley que tipifica y sanciona la violencia digital (Boletín N.º 13.928-07), Informe de Asesoría Técnica Parlamentaria, julio de 2025. Disponible en: https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/37461/2/BCN_violencia_digital_julio_2025.pdf, consultado el 5 de septiembre de 2025.

actores institucionales para consolidar la normativa.[53] Si bien aún no se ha promulgado, su discusión parlamentaria refleja el reconocimiento de la violencia digital como una problemática urgente que requiere respuesta específica desde el ámbito penal.

En definitiva, tanto la Ley N° 21.675 como el Boletín N° 13.928-07 representan pasos incipientes hacia la construcción de un marco normativo que reconozca la violencia digital —incluida la ejercida mediante *deepfakes* sexuales— como una vulneración grave de la dignidad y los derechos fundamentales. Sin embargo, la falta de tipificación expresa y la lentitud del proceso legislativo continúan dejando a las víctimas en una situación de desprotección efectiva.

Ley N° 21.719 sobre Protección de Datos Personales

La Ley N° 21.719[54], promulgada el 13 de diciembre de 2024 y con entrada en vigencia prevista para el 1 de diciembre de 2026, representa un salto cualitativo en la regulación chilena de los datos personales: moderniza y reemplaza la antigua Ley N° 19.628, crea una Agencia de Protección de Datos Personales e incorpora principios, derechos y obligaciones que, como explica Jijena (2025: 63), responden directamente a la influencia del Reglamento General de Protección de Datos (GDPR) europeo.

Entre sus aportes más relevantes se encuentran la creación de la Agencia de Protección de Datos Personales, el establecimiento de un catálogo de derechos reforzados para los titulares y la incorporación de un régimen sancionatorio graduado en infracciones leves, graves y muy graves, con multas que pueden alcanzar hasta 20.000 UTM.[55]

---

53 Senado de la República de Chile, Retoman estudio de proyecto que tipifica y sanciona la violencia digital, 17 de junio de 2025. Disponible en: https://www.senado.cl/comunicaciones/noticias/retoman-estudio-de-proyecto-que-tipifica-y-sanciona-la-violencia-digital, consultado el 5 de septiembre de 2025.
54 Ley N° 21.719 de 2024, que regula la protección y el tratamiento de los datos personales y crea la Agencia de Protección de Datos Personales.
55 ONTIER. "Law 21.719: Regulating the Protection of Personal Data and Creating the Personal Data Protection Agency." Disponible en: https://www.ontier.law/en/law-21-719-regulating-the-protection-of-personal-data-and-creating-the-personal-data-protection-agency/?, consultado el 22 de noviembre de 2025.

Un aspecto especialmente relevante de la Ley N° 21.719 es el mecanismo de responsabilidad civil establecido en su artículo 47. Esta disposición obliga al responsable del tratamiento a indemnizar el daño patrimonial y extrapatrimonial causado cuando infrinja los principios del artículo 3° o los deberes de la ley y genere perjuicio al titular. En principio, este catálogo resulta plenamente aplicable a los *deepfakes* no consentidos, ya que su generación y difusión implica un tratamiento ilícito de datos personales —particularmente de imagen, voz o información biométrica— y supone una afectación directa a derechos de personalidad como la honra, la identidad y la vida privada.

Sin embargo, su eficacia práctica presenta zonas grises. Por una parte, la acción civil solo puede ejercerse una vez que exista una resolución sancionatoria firme de la Agencia, lo que puede demorar la reparación en contextos donde el daño es inmediato y expansivo. Por otra, no siempre es claro que el creador o difusor de un *deepfake* pueda ser calificado técnicamente como "responsable del tratamiento", especialmente cuando se trata de usuarios no profesionales; un problema similar al que ya se ha constatado en el derecho europeo respecto de la aplicabilidad del régimen de protección de datos a infractores individuales y no profesionales (Estella, 2025: 285). A ello se suma la dificultad práctica de identificar al autor material, lo que limita la efectividad del mecanismo indemnizatorio.

Aunque Chile aún no registra casos administrativos ni judiciales tramitados bajo la Ley N° 21.719 en materia de *deepfakes*, la experiencia comparada muestra que las autoridades de protección de datos comienzan a enfrentar este fenómeno desde la óptica del tratamiento ilícito de datos personales. Ello es especialmente relevante frente a hechos como los ocurridos en el Colegio de Santiago, pues en Europa ya existen precedentes sancionatorios directamente vinculados a la generación de "ultrasuplantaciones" íntimas de menores. En efecto, la Agencia Española de Protección de Datos (AEPD) ha resuelto dos casos recientes de *deepfakes* sexuales contra adolescentes, ambos consistentes en la creación y difusión de imágenes falsas de desnudos generadas mediante IA, tratándose de un uso no consentido de datos biométricos e imagen que la autoridad calificó como tratamiento ilícito. Uno de ellos —corres-

pondiente al Expediente N.º EXP202503445[56]— ha sido ampliamente difundido por la prensa como "la primera sanción en Europa por un desnudo falso creado con IA",[57] aunque en realidad existe al menos un caso anterior tramitado por la misma autoridad.

A pesar de la gravedad de los hechos, la cuantía de las multas ha sido objeto de críticas debido a que se consideran sanciones insuficientes para desincentivar conductas altamente lesivas y desproporcionadas frente al daño que experimentan las víctimas, especialmente cuando se trata de menores.[58] Estos antecedentes muestran que, aun en sistemas con regulación avanzada, los mecanismos actuales siguen siendo limitados para responder adecuadamente a la particular intensidad del daño que producen las *deepfakes* sexuales.

Este precedente resulta ilustrativo para Chile en un doble sentido. Primero, confirma que los *deepfakes* íntimos constituyen —al menos en la experiencia europea— un supuesto de tratamiento ilícito de datos personales susceptible de sanción administrativa. Segundo, pone de manifiesto que incluso en países con marcos regulatorios más robustos, las herramientas disponibles todavía se encuentran en fase de ajuste, especialmente en lo referido a la proporcionalidad de las sanciones y a la necesidad de mecanismos de reparación adecuados para quienes ven manipulada su identidad mediante IA. Por lo tanto, si bien la Ley Nº 21.719 ofrece un potencial cauce civil para la reparación del daño causado por estos contenidos, parece aconsejable que su futura aplicación —o incluso reformas posteriores— contemplen reglas específicas para el contenido sintético y sanciones efectivamente disuasorias frente a este fenómeno.

---

56 Agencia Española de Protección de Datos. *Resolución Exp. N.º EXP202503445*. Disponible en https://www.aepd.es/documento/ps-00132-2025.pdf, consultado el 24 de noviembre de 2025.

57 "Protección de Datos impone la primera sanción en Europa por un desnudo falso generado con IA", *La Vanguardia*, 6 de noviembre de 2025. Disponible en: https://www.lavanguardia.com/vida/20251106/11236961/proteccion-datos-impone-primera-sancion-europa-desnudo-falso-generado-ia.html, consultado el 24 de noviembre de 2025.

58 Consejo General de la Abogacía Española, "Deepfakes sexuales y protección de datos", *Opinión y Análisis*, 20 de noviembre de 2025. Disponible en https://www.abogacia.es/actualidad/opinion-y-analisis/deepfakes-sexuales-y-proteccion-de-datos/, consultado el 24 de noviembre de 2025.

Proyecto de ley que regula los sistemas de inteligencia artificial (Boletín Nº 16.821-19)

Finalmente, dentro de las iniciativas legislativas en curso, destaca el Proyecto de Ley que regula los sistemas de inteligencia artificial (Boletín N.º 16.821-19),[59] ingresado el 7 de mayo de 2024 y actualmente en segundo trámite constitucional ante el Senado. Se trata del primer esfuerzo sistemático por establecer un marco transversal para el uso, desarrollo y despliegue de sistemas de IA en Chile.

El proyecto adopta un enfoque basado en riesgos, clasificando los sistemas en cuatro categorías: riesgo inaceptable, alto riesgo, riesgo limitado y sin riesgo evidente (artículo 5º del proyecto de ley). Este diseño reproduce, con notoria fidelidad, la arquitectura del *AI Act* europeo, que desde 2024 estructura su regulación en torno al nivel de riesgo asociado a cada sistema. No obstante, la cercanía conceptual con el modelo europeo no implica una equivalencia funcional, especialmente respecto del tratamiento de los *deepfakes*.

Aunque el texto chileno incorpora principios de intervención humana, transparencia, seguridad, diversidad y responsabilidad, entre otros (artículo 1º letra b.), y proyecta la creación de un Consejo Asesor Técnico de IA, su aproximación se concentra en los sistemas y no en los resultados que estos generan. A diferencia del *AI Act*, que en su artículo 50 regula expresamente las obligaciones de transparencia vinculadas a la generación o manipulación de contenidos —incluidos los *deepfakes*—, el proyecto chileno no contiene ninguna disposición específica sobre contenido sintético, etiquetado, marcado, procedencia o manipulación artificial.

En este sentido, si bien ciertos usos de *deepfakes* podrían ser eventualmente reconducidos a las categorías de "riesgo inaceptable" previstas en el artículo 6º —por ejemplo, cuando impliquen manipulación subliminal, explotación de vulnerabilidades o afectación a la honra e integridad sexual—, esta subsunción es parcial y dependiente del caso concreto. El proyecto chileno no los reconoce como un fenómeno tecnológico

59 Mensaje Nº 063-372 de 7 de mayo de 2024 por el que inicia un Proyecto de Ley de Inteligencia Artificial. Disponible en: https://www.camara.cl/verDoc.aspx?prmID=17048&prmTIPO=INICIATIVA, consultado el 24 de noviembre de 2025.

autónomo que requiere obligaciones preventivas de transparencia, sino que los trata como un daño posible derivado de determinados sistemas. De este modo, solo aquellos que generen un perjuicio grave, o que recaigan sobre grupos especialmente vulnerables, podrían ser prohibidos o sancionados, mientras que la gran mayoría de los contenidos sintéticos —incluidos los destinados a desinformación política, fraude, suplantación de identidad o manipulación mediática— quedaría fuera de cualquier exigencia ex ante de identificación o advertencia al usuario.

En definitiva, la regulación chilena no aborda de forma satisfactoria este fenómeno, al menos no en los términos de transparencia estructural que impone el *AI Act*. Allí donde la normativa europea exige un etiquetado claro, visible y permanente de todo contenido generado o alterado mediante IA —salvo contadas excepciones justificadas—, el proyecto chileno guarda silencio. Así, aunque el proyecto reproduce el enfoque basado en riesgos del *AI Act*, no replica su respuesta normativa específica frente a la manipulación audiovisual, dejando a los *deepfakes* regulados solo de manera indirecta y reactiva.

La responsabilidad civil frente a los *deepfakes* y, en particular, el *deepfake* íntimo no consentido

Desde la perspectiva de la responsabilidad civil, los *deepfakes* también plantean problemas y desafíos. Al ser una tecnología basada en IA, su autonomía relativa, la falta de previsibilidad de los resultados y la dificultad para identificar responsables concretos complejizan la atribución de responsabilidad.

En la actualidad, en Chile no existe un régimen especial de responsabilidad civil en esta materia, ni puede aplicarse la regla de responsabilidad por el hecho de las cosas a los sistemas de IA. Por ello, corresponde evaluar si la legislación vigente resulta suficiente para reparar los daños causados por los *deepfakes* en general, y por los *deepfakes* íntimos no consentidos en particular. Para ello, resulta necesario examinar las dificultades que enfrenta la víctima al momento de acreditar los requisitos clásicos de la responsabilidad civil: la voluntariedad de la acción, el

estándar de diligencia exigible, la relación de causalidad y la extensión del daño indemnizable.[60]

Voluntariedad de la acción

El primer presupuesto de la acción de responsabilidad civil es la voluntariedad de la acción, en cuanto el juicio normativo consiste en imputar a una persona la obligación de reparar el daño causado. El problema radica en que el sistema de inteligencia artificial utilizado para generar el *deepfake* es, en sí mismo, quien produce el resultado lesivo; sin embargo, carece de personalidad jurídica y no puede ser considerado sujeto de derecho. Por ello, la atribución de responsabilidad debe recaer necesariamente en las personas que crean, difunden o se sirven de esa tecnología.

En este contexto resultan relevantes las recomendaciones internacionales sobre ética de la IA, como las de la OCDE y la UNESCO. La primera subraya la necesidad de que los actores de IA respeten el Estado de Derecho, los derechos humanos y aseguren transparencia en el uso de estas herramientas.[61] Por su parte, la UNESCO (2021)[62] ha sostenido que siempre debe existir supervisión humana y que las decisiones generadas por un sistema de IA son imputables a quienes lo diseñan, controlan o utilizan, respondiendo por los daños que surjan de la falta de control en su ciclo de vida. El principio de control humano en su comprensión de responsabilidad ha sido descrito por Bustos del siguiente modo:

> implica asumir que la decisión, recomendación o predicción que emana de un sistema de IA debe ser siempre atribuible a los

60 Seminario "Responsabilidad Civil e Inteligencia Artificial", Magíster en Derecho con mención en Derecho Privado, Facultad de Derecho, Universidad de Chile, primer semestre de 2024. Capítulo elaborado en base a los apuntes de clases y material bibliográfico del curso.
61 Organización para la Cooperación y el Desarrollo Económicos (OCDE), *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 21 de mayo de 2019. Disponible en: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449, consultado el 18 de julio de 2024.
62 UNESCO, Recomendación sobre la ética de la inteligencia artificial, 23 de noviembre de 2021. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa, consultado el 18 de julio de 2024.

> actores de la IA que han debido controlarla o que han decidido aceptar el resultado de la decisión o acción de esta, conforme a la función que tengan en el ciclo de vida del sistema de IA y, por lo tanto, corresponde a los actores de la IA responder por los daños que se causen a las personas y que surjan como resultado de la falta de control sobre el funcionamiento de la IA en todo su ciclo de vida o de aceptar la decisión, recomendación o predicción basada de alguna forma en un sistema de IA (2024: 46).

Estas consideraciones adquieren especial relevancia cuando se trata de *deepfakes* íntimos no consentidos, donde la atribución de responsabilidad no puede quedar diluida en la supuesta "autonomía" de la tecnología. En el caso del Colegio de Santiago, aunque los algoritmos generaron el contenido, fue la decisión consciente de los menores de edad lo que desencadenó el daño, lo cual permite imputarles —y eventualmente a quienes ejercen su cuidado— la responsabilidad civil correspondiente.

En definitiva, aun cuando la creación del *deepfake* se materialice a través de un sistema automatizado, ello no elimina la voluntariedad exigida por la responsabilidad civil, pues siempre existen actores humanos que adoptan la decisión de usar esas plataformas, aceptar sus resultados y difundir el material generado.

En consecuencia, el desafío radica en no aceptar que la ausencia de regulación específica sobre IA constituya un vacío eximente de responsabilidad, sino en aplicar los principios generales del derecho y las recomendaciones internacionales como criterios interpretativos que permitan atribuir la obligación de reparar el daño a los sujetos involucrados. La pregunta clave es si resulta posible incorporar los principios de la OCDE y de la UNESCO a nuestro ordenamiento —ya sea vía equidad, como principios generales de derecho, o como valores presentes en la comunidad jurídica— de modo que puedan ser utilizados para interpretar e integrar las lagunas legales en esta materia. En definitiva, puede sostenerse que

> la falta de una regulación específica de la responsabilidad civil por los daños causados por sistemas de IA no constituye excusa para excluir que los actos u omisiones perjudiciales a

> las personas y que provengan de las decisiones, recomendaciones o predicciones de un sistema de IA puedan ser atribuidos al sujeto en particular (Bustos, 2024: 47).

Estándar de diligencia exigible

En relación con el estándar de diligencia exigible, el punto de partida es determinar si corresponde aplicar un régimen de responsabilidad por culpa —que es la regla general en nuestro derecho civil— o un régimen objetivo. En principio, la víctima debe probar la culpa o dolo de quien crea, utiliza o difunde un *deepfake* sexual no consentido.

El primer desafío surge porque la ilicitud depende de la infracción a un deber de cuidado, definido a partir de lo que una persona razonable y diligente habría hecho en circunstancias similares. Como explica Barros (2010:89), esto supone valorar si el agente podía prever las consecuencias inmediatas de su conducta. Sin embargo, los *deepfakes* basados en redes neuronales profundas (GANs) presentan un grado de imprevisibilidad: aunque un modelo sea entrenado por un programador, el sistema puede generar resultados no anticipados, lo que dificulta precisar hasta qué punto el creador, el usuario o el proveedor podían prever los daños.

Todo lo anterior evidencia la gran dificultad de precisar cuál es la conducta (diligencia) exigible al creador, proveedor y al usuario de un *deepfake*, en cuanto al deber de mitigar o suprimir los riesgos derivados de esta tecnología, en la medida en que dichos riesgos muchas veces son inciertos o desconocidos.

La cuestión se torna especialmente crítica en el caso de los *deepfakes* sexuales, ya que el peligro de vulneración de derechos fundamentales como la honra, la intimidad y la integridad psíquica resulta evidente. Incluso si el sistema presenta cierto grado de autonomía, la decisión humana de activar, utilizar y difundir esta tecnología con fines sexuales no consentidos constituye una conducta claramente negligente o dolosa, por cuanto presupone la previsibilidad de los daños que tales actos pueden ocasionar.

Un segundo problema está en identificar qué actores deben responder dentro de la cadena de producción y difusión: ¿el usuario que ingresa

las instrucciones al sistema?; ¿el creador de la tecnología?; ¿el proveedor que facilita el acceso a la plataforma?; ¿o las redes sociales que permiten la masificación del contenido? En este sentido, es posible aplicar el sistema general de responsabilidad por culpa, pero es importante fortalecer los deberes de diligencia diferenciados para cada uno (seguridad, información, control, supervisión), de modo que se pueda precisar la carga que corresponde a cada interviniente sin necesidad de reformar la ley.

No obstante, la carga probatoria para las víctimas suele ser excesiva, especialmente en casos complejos donde intervienen múltiples actores y cuando se trata de mujeres, niñas o adolescentes expuestas a *deepfakes* sexuales. Por ello, cabe debatir la introducción de presunciones de culpa o regímenes de responsabilidad estricta. Esta opción encuentra sustento en la protección de las víctimas en situación de desventaja estratégica y en la calificación del *deepfake* sexual como una actividad de riesgo inaceptable, siguiendo la clasificación de la *IA Act* europea.

Así, podría avanzarse hacia el establecimiento de un régimen especial que contemple un sistema escalonado de responsabilidad: aplicar la regla general de culpa para los *deepfakes* de riesgo mínimo o nulo, establecer presunciones de culpa en casos de alto riesgo, y reconocer un régimen de responsabilidad estricta para los *deepfakes* sexuales no consentidos, en atención a su especial peligrosidad y a la magnitud de los daños que generan.

Causalidad

El requisito de causalidad debe acreditarse tanto bajo el régimen general de responsabilidad por culpa como en eventuales supuestos de responsabilidad estricta. En cualquier caso, el daño debe ser consecuencia directa y necesaria de un hecho atribuible a un sujeto determinado.

En materia de *deepfakes*, acreditar el nexo causal puede resultar particularmente complejo. Se trata de una tecnología que funciona mediante redes neuronales profundas, en la cual intervienen múltiples actores en distintas fases —creación, utilización y difusión—, muchas veces bajo anonimato o amparados en grandes plataformas multinacionales

como Google o Amazon. La facilidad de circulación y replicación en línea incrementa aún más la dificultad de precisar el origen del *deepfake* y de identificar al sujeto que lo generó. Por ello, en abstracto, surge la necesidad de determinar qué deber de conducta corresponde a cada interviniente y cuál es el nivel de diligencia exigible.

Ahora bien, en el caso ejemplificador mencionado —el escándalo del Colegio en Santiago— el problema de la causalidad no reviste la complejidad descrita. En este supuesto fue posible identificar a los alumnos responsables de la creación y difusión de los *deepfakes* sexuales, de modo que las reglas generales de la responsabilidad civil por culpa resultan plenamente aplicables. La imputación del daño se dirige directamente a los autores, sin necesidad de recurrir a esquemas especiales de atribución.

Ello no impide advertir que, en otros casos de *deepfakes* íntimos no consentidos —en particular cuando los responsables no son fácilmente identificables o cuando intervienen plataformas tecnológicas globales— pueden presentarse serias dificultades para trazar el nexo causal y atribuir jurídicamente el daño. En tales situaciones, cobra relevancia precisar la función de cada agente interviniente y reconducir la imputación al cumplimiento del estándar de cuidado exigible. Para ello, resultan útiles los principios éticos de la OCDE y las recomendaciones de la UNESCO, que enfatizan la necesidad de transparencia, trazabilidad y supervisión humana en el desarrollo y uso de la inteligencia artificial.

Extensión del daño indemnizable

El último elemento a considerar es la extensión del daño indemnizable, cuestión que suele presentarse como una de las mayores dificultades en la responsabilidad civil derivada del uso de inteligencia artificial, incluidos los *deepfakes*.

En términos generales, la autonomía e imprevisibilidad de estos sistemas dificulta precisar cuál es la conducta exigible a cada uno de los intervinientes para mitigar o suprimir los riesgos que generan. Surge entonces la interrogante sobre la eventual responsabilidad no solo de quienes hacen un uso directo de la tecnología, sino también de los desa-

rrolladores o difusores de herramientas de IA, en la medida en que su puesta en circulación posibilita la producción de imágenes adulteradas. Sin embargo, cuando se trata de los usuarios inmediatos —como en el caso de los alumnos del Colegio de Santiago— el estándar de conducta resulta claro: cualquier persona puede prever que la creación y difusión de imágenes desnudas de otra conlleva un daño extrapatrimonial evidente. En este escenario, las reglas generales de la culpa bastan para fundamentar la responsabilidad.

Ahora bien, en un plano más abstracto, la dificultad radica en determinar cómo se extiende la responsabilidad en casos donde la previsibilidad no es evidente y el daño proviene de riesgos inciertos o imprevistos propios de la tecnología. En este contexto suele invocarse el principio precautorio, que impone la adopción de medidas destinadas a evitar la materialización de tales riesgos. Sin embargo, su aplicación tensiona la noción clásica de culpa —que exige previsibilidad y certeza— y tampoco se ajusta plenamente a un régimen de responsabilidad objetiva, pues incluso allí debe demostrarse la causalidad. Como advierte Banfi (2019: 655), el principio precautorio "transtorna los conceptos de daño, culpa y nexo causal" y, si se aplica sin cautela, puede derivar en "una responsabilidad estricta absoluta, donde ni siquiera el caso fortuito le permitiría exonerarse del deber de resarcir los perjuicios emanados de su actividad".

En consecuencia, mientras que en el caso del Colegio de Santiago la extensión del daño indemnizable no ofrecía mayores dificultades —ya que los perjuicios eran perfectamente previsibles—, en otros escenarios de *deepfakes* la problemática persiste. Los riesgos de daños inciertos no parecen ser indemnizables ni bajo la causalidad fáctica ni bajo las teorías de imputación objetiva, que siempre exigen conexión con el curso normal de los hechos. Esto evidencia la necesidad de explorar respuestas alternativas para dar cuenta de daños originados en tecnologías autónomas y difícilmente controlables.

Una primera aproximación proviene de las recomendaciones de la UNESCO sobre ética de la IA y de los principios de la OCDE, que destacan la supervisión humana y la transparencia como herramientas de gestión de riesgos. Estos instrumentos reconocen la autonomía y la imprevisibilidad de los sistemas de IA, y promueven un enfoque

precautorio en su desarrollo y uso. Un principio especialmente útil es el de seguridad y protección, por el cual

> los daños no deseados (riesgos de seguridad) y las vulnerabilidades a los ataques (riesgo de protección) deberían ser evitados y deberían tenerse en cuenta, prevenirse y eliminarse a lo largo del ciclo de vida de los sistemas de IA para garantizar la seguridad y protección de seres humanos, del medio ambiente y de los ecosistemas.[63]

Este principio ordena implementar acciones anticipadas necesarias para evitar que se materialicen ciertos riesgos o daños que resultan previsibles.

Otra alternativa la ofrece la Resolución N° 53 del Parlamento Europeo (2017), que propone un sistema de gestión de riesgos que "no se centra en la persona «que actuó de manera negligente» como personalmente responsable, sino en la persona que es capaz, en determinadas circunstancias, de minimizar los riesgos y gestionar el impacto negativo".[64] A ello se suma el *AI Act*, que introduce un enfoque regulatorio basado en niveles de riesgo. Según este modelo, lo determinante ya no es la previsibilidad del daño, sino la omisión de medidas para evitarlo o reducirlo.

En Chile, esta aproximación comienza a reflejarse en un proyecto de ley presentado el 7 de mayo de 2024, que clasifica los sistemas de IA en niveles de riesgo —inaceptable, alto, limitado y mínimo— siguiendo el modelo europeo. Aplicado a los *deepfakes*, este sistema permitiría prohibir expresamente aquellos de carácter sexual no consentido que explotan la vulnerabilidad de mujeres, niños, niñas y adolescentes, como en el caso de estudio.

---

63 UNESCO. Recomendación sobre la ética de la inteligencia artificial, 23 de noviembre de 2021, cap. III.2, principio n.° 27. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000373434_spa, consultado el 5 de septiembre de 2025.

64 Parlamento Europeo. Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)), Diario Oficial de la Unión Europea, apdo. 55 ("Responsabilidad"). Disponible en: https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52017IP0051&from=EN, consultado el 5 de septiembre de 2025.

Conclusiones

Este trabajo ha mostrado que los *deepfakes* representan un desafío transversal para los ordenamientos jurídicos, pues no solo cuestionan la autenticidad de la información y la confianza pública, sino que además amenazan de manera directa derechos fundamentales. La afectación es particularmente grave en el ámbito de la violencia de género, ya que las mujeres constituyen el principal grupo expuesto al fenómeno del *deepfake* íntimo no consentido, expresión contemporánea de la violencia digital.

El análisis comparado evidencia que otras jurisdicciones han avanzado con mayor decisión. La Unión Europea, mediante el *AI Act*, estableció un régimen de clasificación por riesgos que prohíbe los *deepfakes* de riesgo inaceptable; Australia ha dictado normas específicas en materia de violencia digital, y diversos estados en EE. UU. han tipificado expresamente la creación y difusión de *deepfakes* sexuales no consentidos. En contraste, Chile carece de una regulación específica, lo que quedó patente en el caso del Colegio de Santiago, donde menores fueron víctimas de falsificaciones pornográficas. Este episodio cuestionó la suficiencia de los remedios legales que contempla nuestro ordenamiento jurídico frente a los daños que provoca esta tecnología.

Si bien el país ha avanzado en la construcción de un marco normativo en materia de violencia de género y violencia digital —especialmente con la promulgación de la Ley N° 21.675 y la tramitación del Boletín N° 13.928-07—, estos desarrollos aún resultan insuficientes para enfrentar de manera integral los riesgos que genera la inteligencia artificial en la producción de contenidos falsificados. Como demuestra el análisis del caso chileno, la respuesta actual del ordenamiento es fragmentaria y reactiva: el recurso de protección permite solo una tutela *ex post* frente a vulneraciones ya consumadas; el derecho penal carece de tipos que aborden específicamente la creación sintética de imágenes íntimas —incluidos los artículos 161-C y 161-D, cuya estructura típica exige la existencia de un registro real—; la Ley N° 21.719 ofrece un cauce relevante desde la protección de datos, pero con importantes límites prácticos y procedimentales; y el proyecto de ley sobre inteligencia artificial

replica el enfoque de riesgo del *AI Act*, sin incorporar obligaciones de transparencia o etiquetado de contenido sintético. En consecuencia, Chile aún no cuenta con un marco coherente que permita prevenir, identificar y sancionar efectivamente los *deepfakes*, lo que evidencia la necesidad de avanzar hacia regulaciones específicas que aborden los desafíos estructurales que plantea la manipulación audiovisual mediante IA, desde una perspectiva de derechos fundamentales y con especial atención a su dimensión de género.

En el ámbito civil, el examen de la voluntariedad, del estándar de diligencia exigible, de la causalidad y de la extensión del daño indemnizable evidencia que las reglas generales de la responsabilidad civil ofrecen respuestas suficientes en casos claros, como el ocurrido en el Colegio de Santiago. Sin embargo, en escenarios más difusos, donde intervienen múltiples actores —creadores, usuarios, proveedores y plataformas—, la autonomía relativa de los sistemas de IA, su imprevisibilidad y la circulación masiva de los contenidos generan serias dificultades para imputar responsabilidad conforme a los presupuestos clásicos de la culpa.

Frente a estas tensiones, no debe admitirse que la ausencia de regulación específica opere como un vacío eximente de responsabilidad. Corresponde, en cambio, aplicar los principios generales del derecho y utilizar como criterios interpretativos las recomendaciones internacionales de la UNESCO y de la OCDE, que insisten en la centralidad del control humano, la transparencia y la trazabilidad. A su vez, es preciso fortalecer el catálogo de deberes de diligencia de todos los actores intervinientes en el ciclo de vida de la IA, estableciendo estándares diferenciados según su posición en la cadena tecnológica.

Por otra parte, resulta deseable que el ordenamiento jurídico chileno avance hacia un sistema de responsabilidad especial en esta materia, fundado en un enfoque de riesgos. Dicho sistema debiera mantener la regla de la culpa en los supuestos de riesgo mínimo, pero incorporar presunciones de culpa en escenarios de alto riesgo y un régimen de responsabilidad estricta en casos de riesgo inaceptable, dentro de los cuales debe comprenderse el *deepfake* íntimo no consentido. Esta solución permitiría equilibrar la protección de las víctimas —frecuentemente mujeres, niñas y adolescentes en situación de especial vulnerabilidad—

con la necesidad de incentivar que todos los actores intervinientes en el ciclo de vida de la IA adopten estándares de diligencia acordes con su posición.

Referencias

Banfi, Cristián. (2019). "Riesgos en la aplicación del principio precautorio en responsabilidad civil y ambiental". *Revista Chilena de Derecho*, 46 (3): 643–667. Disponible en https://www.scielo.cl/pdf/rchilder/v46n3/0718-3437-rchilder-46-03-643.pdf.

Barros, Enrique (2010). *Tratado de responsabilidad extracontractual*. 1ª ed. Santiago: Editorial Jurídica de Chile.

Bustos, Magdalena (2024). "El principio ético de la intervención humana en la responsabilidad civil". *Acta Bioethica*, 30 (1), 41–49. Disponible en https://www.scielo.cl/pdf/abioeth/v30n1/1726-569X-abioeth-30-01-41.pdf.

Chesney, Robert y Danielle Citron (2019). "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security". *California Law Review*, 107 (6): 1753-1820. DOI: https://doi.org/10.2139/ssrn.3213954.

Estella, Antonio (2025). "La regulación de las deepfakes (ultrasuplantaciones) en el reglamento de la UE sobre Inteligencia Artificial". *Revista de Administración Pública*, 226: 261-290. DOI: https://doi.org/10.18042/cepc/rap.226.11

Fletcher, John. (2018). "Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance". *Theatre Journal*, 70 (4): 455-471. DOI: https://doi.org/10.1353/tj.2018.0097.

Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville y Yoshua Bengio (2014). "Generative Adversarial Nets". *Advances in Neural Information Processing Systems*, 27: 2672–2680. Disponible en https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

Jijena, Renato. (2025). *Responsables de datos personales. Análisis Leyes N.º 19.628, 21.719 y el RGPD.* Valencia: Tirant lo Blanch.

Simó, Elisa (2023). "Retos jurídicos derivados de la Inteligencia Artificial Generativa. Deepfakes y violencia contra las mujeres como supuesto de hecho". *InDret*, 2: 493–515. DOI: https://doi.org/10.31009/InDret.2023.i2.11

Tolosana, Rubén, Rubén Vera-Rodríguez, Julián Fierrez, Aythami Morales, y Javier Ortega-García (2020). "Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection". *Information Fusion*, 64: 131-148. DOI: https://doi.org/10.1016/j.inffus.2020.06.014.

Wagner, Travis L. y Ashley Blewer (2019). "The Word Real Is No Longer Real: Deepfakes, Gender, and the Challenges of AI-Altered Video". *Open Information Science*, 3 (1): 32-46. Disponible en https://www.degruyterbrill.com/document/doi/10.1515/opis-2019-0003/html?lang=en.

Westerlund, Mika. (2019). "The Emergence of Deepfake Technology: A Review". *Technology Innovation Management Review*, 9 (11): 40-53. DOI: https://doi.org/10.22215/timreview/1282.

Sobre la autora

Vanesa Jabbaz Rosenbaum es abogada y licenciada en Ciencias Jurídicas y Sociales por la Universidad de Chile. Además, es candidata a Magíster en Derecho con mención en Derecho Privado por la misma institución. Es profesora instructora del Departamento de Derecho Privado de la Universidad de Chile.

vjabbaz@derecho.uchile.cl
https://orcid.org/0000-0002-5383-977X