

JUSTICIA EN LA ERA DEL ENGAÑO DIGITAL: *DEEPPFAKES* Y LA AMENAZA A LA PRUEBA PENAL

*Carlos René Galdames Gaytán**

“Aquel que sabe fingir confusión
y usar el engaño vencerá”.

Sun Tzu

RESUMEN

Este artículo analiza los desafíos jurídicos y tecnológicos que los *deepfakes* plantean para el sistema de justicia penal chileno. Se examina su impacto en los derechos fundamentales, la admisibilidad probatoria y la necesidad de reformas normativas y técnicas. El estudio propone medidas integrales que incluyen la tipificación penal de conductas relacionadas, la adopción de estándares de autenticación digital y la formación continua de los operadores jurídicos, con el fin de garantizar la seguridad jurídica y la protección de los derechos humanos en la era digital.

Palabras clave: *Deepfakes*, prueba penal, derechos fundamentales, autenticación digital, reforma normativa.

INTRODUCCIÓN

El vertiginoso avance de la inteligencia artificial ha posibilitado la creación de *deepfakes*, que son contenidos audiovisuales hiperrealistas que reproducen rostros, voces y gestos con tal precisión que resulta difícil, e incluso imposible, diferenciarlos de grabaciones auténticas.

Esta tecnología plantea serias amenazas sociales y jurídicas, especialmente en el ámbito penal, amenazando derechos fundamentales como la presunción de inocencia, el derecho a un juicio justo y la legalidad de la prueba, facilitando incorporar archivos falsificados que pueden distorsionar la percepción del juez.

* Alumno del Programa de Derecho Vespertino de la USS.
Correo electrónico: crgg1972@gmail.com

Además pueden afectar derechos personalísimos, como el honor, la imagen y la vida privada, por medio de montajes difamatorios o sexuales. En Chile, ni la Ley N° 21.459 ni los mecanismos probatorios actuales ofrecen una respuesta suficiente ante estos riesgos.

Por ello, se requiere una reforma normativa integral –sustantiva, procesal y probatoria– que incorpore estándares técnicos de autenticación digital, garantizando la protección de los derechos fundamentales.

Contextualizado el problema, surge la interrogante: ¿Cómo debe responder el ordenamiento jurídico chileno frente a la creciente amenaza que representan los *deepfakes* para la integridad de la prueba penal?

CONCEPTUALIZACIÓN Y NATURALEZA TÉCNICA DE LOS *DEEPPFAKES*

Los *deepfakes* son contenidos audiovisuales –principalmente videos y audios– creados mediante inteligencia artificial, que simulan con alto realismo rostros, gestos, voces y movimientos humanos.

Técnicamente, se elaboran utilizando redes neuronales profundas, en particular algoritmos de aprendizaje automático como las Generative Adversarial Networks (GANs), que enfrentan dos modelos: uno genera contenido falso y el otro lo evalúa, mejorándolo progresivamente. Esto da lugar a archivos que resultan casi indistinguibles de los reales, tanto para el ojo humano como para algunos sistemas computacionales.

Desde el ámbito jurídico, los *deepfakes* representan una forma novedosa de manipulación digital que desafía los principios tradicionales del Derecho Penal y Civil. Su uso indebido puede constituir delitos, como suplantación de identidad, calumnias, injurias, extorsión y vulneraciones a derechos personalísimos, como el honor, la imagen y la vida privada, además de comprometer la valoración de la prueba penal.

Su sofisticación técnica agrava el problema: las imágenes son coherentes, los movimientos faciales se sincronizan con el discurso y las voces imitan patrones fonéticos únicos; esta verosimilitud extrema dificulta su detección, incluso con herramientas profesionales, facilitando su uso como prueba falsa o su difusión con fines delictivos. Por ello, comprender esta complejidad y su regulación jurídica resulta clave para enfrentar los riesgos y diseñar una respuesta adecuada.

IMPACTOS JURÍDICOS Y AMENAZAS EN EL SISTEMA DE JUSTICIA

La irrupción de los *deepfakes* en el ecosistema digital plantea graves riesgos para el sistema jurídico, especialmente en lo penal. Uno de los efectos más relevantes se manifiesta en la formación de prueba: al posibilitar la creación de registros audiovisuales falsos pero convincentes, los *deepfakes* debilitan el principio de presunción de inocencia.

Una imagen o video manipulado puede ser suficiente para imputar falsamente a una persona, invirtiendo ilegítimamente la carga probatoria y afectando el criterio del tribunal.

En el ámbito procesal, surge un desafío crucial respecto de la admisibilidad y valoración de pruebas, ya que la falta de mecanismos generalizados y fiables para autenticar el origen y la integridad de archivos digitales compromete la seguridad jurídica.

La duda acerca de la autenticidad de una prueba audiovisual puede derivar en su exclusión, la nulidad del procedimiento o en decisiones judiciales erróneas basadas en evidencia adulterada.

Desde el Derecho Civil, el uso de *deepfakes* con fines difamatorios, extorsivos o de suplantación implica, como se ha dicho, responsabilidad por vulneración de derechos personalísimos, como el honor, la imagen y la vida privada; la rápida propagación de estos contenidos puede generar un daño reputacional irreparable, agravado por la dificultad de identificar al autor y lograr la remoción de esos contenidos de las redes.

En definitiva, los *deepfakes* amenazan pilares fundamentales del sistema de justicia, exigiendo revisar profundamente los marcos normativos actuales.

VULNERACIÓN DE DERECHOS FUNDAMENTALES

El uso malicioso de *deepfakes* constituye una amenaza directa a diversos derechos fundamentales que sostienen el Estado de Derecho.

En primer lugar, se vulneran los derechos al honor, a la imagen y a la vida privada cuando estos montajes se emplean para generar contenidos difamatorios, sexuales o humillantes; la reproducción no consentida de la identidad de una persona, aun sin participación en la grabación original, afecta su dignidad y reputación, provocando daños emocionales y sociales muchas veces irreparables.

En el ámbito judicial, los *deepfakes* comprometen gravemente el derecho al debido proceso y al juicio justo; la posibilidad de introducir pruebas audiovisuales manipuladas puede inducir al error a jueces, fiscales y defensores, debilitando garantías procesales esenciales.

También pueden distorsionar la reconstrucción de los hechos y afectar la imparcialidad del procedimiento al generar prejuicios contra el imputado o la víctima; los riesgos de suplantación e incriminación indebida son alarmantes, ya que una persona puede ser falsamente implicada en actos delictivos o atribuírsele conductas que nunca realizó.

En víctimas de delitos sexuales o violencia de sexo, los *deepfakes* pueden causar nuevas formas de revictimización, ampliando el daño original. Así, esta tecnología erosiona derechos esenciales y exige medidas legales, tecnológicas y educativas urgentes que protejan a las personas ante esta agresión digital.

ANÁLISIS DE JURISPRUDENCIA NACIONAL E INTERNACIONAL

En Chile, la jurisprudencia pertinente a *deepfakes* aún es incipiente o prácticamente inexistente, lo que revela un desfase entre el acelerado desarrollo tecnológico y la capacidad del ordenamiento jurídico para responder oportunamente a sus implicancias.

La falta de doctrina consolidada y de precedentes judiciales claros deja a los tribunales sin directrices específicas para abordar casos donde la prueba audiovisual esté manipulada, lo que genera un vacío normativo y una preocupante inseguridad jurídica.

Sin embargo, existe un caso afín relevante: la Corte Suprema, en el Rol N.º 71.329-2021, condenó la creación de un perfil falso en redes sociales que suplantaba a un tercero y publicaba contenido denigrante que afectaba su reputación.

Aunque no se trató de un *deepfake* propiamente tal, el fallo reconoció la vulneración de derechos fundamentales como la honra y la vida privada (CPR. Art. 19 N° 4), estableciendo criterios respecto del uso malicioso de tecnologías digitales para dañar la imagen de una persona.

Por su parte, la Ley N° 21.459 acerca de delitos informáticos (2022) sanciona conductas como el acceso indebido a sistemas o la interceptación de comunicaciones, pero aún no contempla expresamente la creación o difusión de *deepfakes*, lo que ha motivado propuestas doctrinales para su futura incorporación.

En contraste, la experiencia internacional ha comenzado a establecer criterios útiles. En la Unión Europea, el Reglamento General de Protección de Datos (GDPR) fue invocado por el Tribunal Regional de Frankfurt (Alemania, 2020), que ordenó indemnizar a una víctima cuya imagen fue usada en un *deepfake* sexual sin consentimiento, por violación del artículo 6 del GDPR y del derecho a la dignidad.

En el Reino Unido, en el caso *Persons Unknown* (High Court, 2022), una mujer obtuvo una orden para retirar un *deepfake* pornográfico, invocando el *tort of misuse of private information* y el UK GDPR.

En Estados Unidos, casos como *Doe v. Meta Platforms Inc.* (California, 2023) han permitido demandas civiles por *deepfakes* no consentidos, fundadas en suplantación de identidad, difamación y negligencia; además, se ha reconocido la necesidad de elevar los estándares probatorios para admitir evidencia audiovisual.

LIMITACIONES TECNOLÓGICAS Y DEL ENTORNO DIGITAL

Uno de los principales desafíos que enfrenta actualmente el sistema de justicia ante los *deepfakes* es la escasa disponibilidad de herramientas tecnológicas eficaces y generalizadas para su verificación y autenticación.

Aunque existen desarrollos experimentales en el ámbito académico y corporativo —como *software* basado en inteligencia artificial inversa o el uso de *blockchain* para

trazabilidad digital–, estas soluciones aún no se han integrado de forma sistemática en los procesos judiciales ni están al alcance cotidiano de fiscales, defensores o jueces.

La falta de estándares técnicos homologados y accesibles dificulta la detección precisa de contenidos manipulados. Como resultado, el sistema judicial no siempre dispone de medios confiables para determinar si una prueba audiovisual ha sido alterada.

Esta debilidad se vuelve crítica en casos donde la prueba digital es central, como delitos sexuales grabados, imágenes de cámaras de vigilancia o declaraciones registradas.

Además, la capacidad humana para identificar *deepfakes* es limitada; el alto realismo de estos montajes puede superar la percepción visual o auditiva de jueces y operadores jurídicos, incluso cuando actúan con buena fe; en este contexto, la ausencia de herramientas forenses adecuadas no solo cuestiona la autenticidad de la prueba, sino que también compromete gravemente la seguridad jurídica del proceso judicial.

PROPUESTA INTEGRAL DE RESPUESTA Y REFLEXIÓN FINAL

Frente a los riesgos que los *deepfakes* representan para el proceso penal y los derechos fundamentales, se requiere una respuesta estructural e interdisciplinaria que articule reformas legales, tecnológicas, institucionales y educativas.

En el plano normativo, urge una reforma integral que abarque: (i) modificaciones sustantivas para tipificar conductas lesivas vinculadas a la creación y difusión de *deepfakes*; (ii) reformas procesales que establezcan protocolos claros para su incorporación, valoración o exclusión como prueba, y (iii) una actualización del estándar probatorio, exigiendo requisitos técnicos mínimos para admitir evidencia audiovisual generada o puesta a circular en entornos digitales.

Asimismo, resulta indispensable la implementación de estándares técnicos de autenticación digital, que permitan verificar la integridad y el origen del material audiovisual; estos deben incorporar marcas de agua digitales, sistemas de metadatos certificados, mecanismos de trazabilidad y verificación criptográfica.

Su desarrollo y mantención debe ser respaldado por entidades públicas y privadas, garantizando interoperabilidad y acceso universal para jueces, fiscales y defensores.

Un eje esencial de esta estrategia es la formación continua de operadores jurídicos en tecnologías emergentes, especialmente en generación y detección de *deepfakes*, fortaleciendo así sus competencias digitales y evitando errores interpretativos que comprometan decisiones ajustadas a derecho.

Por último, la educación digital de la ciudadanía cumple un rol preventivo clave; concientizar acerca de los riesgos, derechos vulnerables y vías de denuncia, empodera a las personas y fomenta una cultura jurídica acorde con los desafíos de la era digital.

CONCLUSIONES GENERALES

El fenómeno de los *deepfakes* desafía las bases mismas del sistema de justicia penal al comprometer la integridad de la prueba, la presunción de inocencia, el debido proceso y los derechos personalísimos; su sofisticación técnica y capacidad de engaño superan los marcos normativos y tecnológicos vigentes, generando vacíos legales, inseguridad jurídica y nuevas formas de daño.

Chile enfrenta una creciente exposición a estos riesgos sin una legislación específica, ni jurisprudencia consolidada que permita abordarlos adecuadamente.

La respuesta debe ser integral: requiere actualizar el Derecho sustantivo y procesal, incorporar herramientas tecnológicas forenses, formar operadores jurídicos y promover una ciudadanía digital crítica; solo así será posible preservar los principios fundamentales del proceso penal en un contexto de acelerada transformación tecnológica.

REFLEXIÓN FINAL

El desafío consiste en equilibrar la innovación con la protección de los derechos humanos, evitando que el progreso digital se convierta en una regresión jurídica. Por ello, es imperativo visibilizar esta problemática y abrir el debate doctrinario y práctico, permitiendo que –desde el Derecho chileno– se construyan respuestas acordes con los desafíos de la era digital.

REFERENCIAS

- CHESNEY, ROBERT, Y DANIELLE CITRON (2019). “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security”. *California Law Review* 107, N° 6: 1753-1819.
- CHILE. LEY N° 21.459 (2022). Establece normas sobre delitos informáticos y deroga la Ley N° 19.223. *Diario Oficial de la República de Chile*, 20 de junio de 2022.
- SEON (2025). “¿Qué son los *deepfakes*?”. *Glosario SEON*.
- AMAZON WEB SERVICES (AWS) (2025). “What Is a GAN? A Generative Adversarial Network (GAN) Is a Deep Learning Architecture...”. *AWS Documentation*.
- VALENZUELA SALDÍAS, JONATAN (2022). “El manejo masivo de datos y la inclusión probatoria en el proceso penal chileno”. *Revista Chilena de Derecho Procesal*.
- VALENZUELA, JONATAN (26 de julio de 2022). “*Deepfakes* en Chile: cuando la justicia enfrenta la manipulación digital”. *Idealex.press*.
- ACADEMIA JUDICIAL DE CHILE (2024). *Repercusiones de la inteligencia artificial generativa en las amenazas a los derechos fundamentales*. Santiago: Academia Judicial.

CORTE SUPREMA DE CHILE. Rol N° 71.329-2021.

TRIBUNAL REGIONAL DE FRANKFURT (2020). Caso de indemnización por *deepfake* sexual, aplicando art. 6 del GDPR. Véase: Sitio oficial de protección de datos en Alemania.

DOE V. META PLATFORMS INC., California, 2023.

AMBER VIDEO. “Amber Authenticate: Blockchain-Based Video Authentication Technology”. En *Deepfakes* en Chile: cuando la justicia enfrenta la manipulación digital. Idealex.press, 26 de julio de 2022.

ECHEVERRÍA, SILVIO, LUIZ BONOMI Y AVELINO ZORZO (27 de julio de 2018). “B-CoC: A Blockchain-Based Chain of Custody for Evidences Management in Digital Forensics”. arXiv preprint arXiv:1807.10359 [cs.CR].

PERSONS UNKNOWN (High Court, Reino Unido, 2022). Orden judicial por *deepfake* pornográfico. Véase: [UK GDPR y jurisprudencia en misuse of private information].

TEPPERSPECTIVES. “Battling Deepfakes: Is Blockchain the Answer?”. Tepper School of Business, Carnegie Mellon University, 2024.

KORSHUNOV, PAVEL, Y SÉBASTIEN MARCEL (2020). “Deepfakes: A New Threat to Face Recognition? Assessment and Detection”. Journal of Visual Communication and Image Representation 66: art. 102596.

“HUMAN PERFORMANCE IN DETECTING DEEPFAKES: A SYSTEMATIC REVIEW”. Journal of Behavioral and Experimental Finance 41 (2024): art. 100774.

IPROOV. DEEPFAKE BLINDSPOT: ONLY 0.1% OF PEOPLE CAN ACCURATELY DETECT AI-GENERATED DEEPFAKES. Estudio de percepción, febrero de 2025.