

LA CRISIS EPISTÉMICA DE LA PRUEBA AUDIOVISUAL EN LA ERA DE LOS DEEPFAKES: MARCO PROCESAL-NORMATIVO PARA LA AUTENTICACIÓN Y VALORACIÓN JUDICIAL⁵⁵⁶

THE EPISTEMIC CRISIS OF AUDIOVISUAL EVIDENCE IN THE AGE OF DEEPFAKES: A PROCEDURAL-NORMATIVE FRAMEWORK FOR AUTHENTICATION AND JUDICIAL EVALUATION

Fernando Ramos-Zaga

Abogado por la Universidad Peruana de Ciencias Aplicadas (UPC) y magíster en Gerencia Social por la Pontificia Universidad Católica del Perú (PUCP). Especialista en filosofía del derecho, regulación y bioética, desarrolla investigación interdisciplinaria orientada al análisis crítico de las relaciones entre el derecho, las políticas públicas y la ética aplicada. Actualmente, forma parte de la Dirección de Investigación de la Universidad Privada del Norte (UPN), donde contribuye al desarrollo y fortalecimiento de proyectos académicos con alto rigor científico. Email: fernanramos@upn.edu.pe
ORCID: <https://orcid.org/0000-0001-6301-9460>.

RESUMEN: La proliferación de contenidos audiovisuales manipulados mediante inteligencia artificial representa un desafío inédito para la función epistémica del proceso judicial contemporáneo. La presente investigación analiza críticamente los marcos normativos y dogmáticos de autenticación y valoración probatoria en los sistemas del *common law* y del derecho continental, identificando las insuficiencias estructurales que los *deepfakes* revelan en ambas tradiciones jurídicas. Mediante un análisis comparativo trans-sistémico, se examina la erosión de los criterios tradicionales de autenticidad, relevancia y valor probatorio, así como los impactos cognitivos y epistemológicos que la manipulación algorítmica genera en jueces, peritos y

testigos. Se evalúan las estrategias normativas emergentes en la Unión Europea y Estados Unidos, particularmente el Reglamento (UE) 2024/1689 y su potencial incorporación en los mecanismos procesales de control judicial. Como resultado, se propone un modelo híbrido de control integral que articula una fase de autenticación técnico-jurídica basada en trazabilidad algorítmica y cadena de custodia digital, seguida de una etapa de valoración racional fundamentada en la sana crítica judicial ampliada. Este marco procesal busca restablecer la legitimidad epistémica del sistema judicial mediante la integración de criterios de fiabilidad tecnológica y razonabilidad argumentativa, garantizando así la tutela judicial

⁵⁵⁶ Artigo recebido em 01/11/2025 e aprovado em 16/12/2025.

efectiva en contextos de alta complejidad digital.

PALABRAS CLAVE: *deepfakes*, prueba audiovisual, autenticación probatoria, valoración judicial, inteligencia artificial.

ABSTRACT: The proliferation of audiovisual content manipulated through artificial intelligence represents an unprecedented challenge to the epistemic function of contemporary judicial proceedings. This research critically analyzes the normative and dogmatic frameworks for authentication and evaluation of evidence in *common law* and civil law systems, identifying the structural inadequacies that *deepfakes* reveal in both legal traditions. Through a trans-systemic comparative analysis, the erosion of traditional criteria of authenticity, relevance, and probative value is examined, alongside the cognitive and epistemological impacts that algorithmic manipulation generates on judges, expert witnesses, and testimonial evidence. Emerging regulatory strategies in the European Union and the United States are evaluated, particularly Regulation (EU) 2024/1689 and its potential incorporation into procedural mechanisms of judicial control. As a result, a hybrid model of comprehensive control is proposed, articulating a phase of technical-legal authentication based on algorithmic traceability and digital chain of custody, followed by a stage of rational evaluation grounded in expanded judicial sound criticism. This

procedural framework seeks to reestablish the epistemic legitimacy of the judicial system through the integration of technological reliability criteria and argumentative reasonableness, thereby guaranteeing effective judicial protection in contexts of high digital complexity.

KEYWORDS: *deepfakes*, audiovisual evidence, evidentiary authentication, judicial evaluation, artificial intelligence.

INTRODUCCIÓN

La relación entre verdad procesal y medios de prueba constituye uno de los pilares fundamentales sobre los que se asienta la legitimidad del sistema judicial contemporáneo. Durante décadas, la evidencia audiovisual ha operado como instrumento privilegiado para la reconstrucción de los hechos controvertidos, beneficiándose de una presunción de objetividad que derivaba tanto de su naturaleza técnica como de su aparente inmunidad frente a la manipulación humana deliberada. Sin embargo, el advenimiento de tecnologías de síntesis audiovisual basadas en redes neuronales generativas adversarias ha fracturado esta confianza epistémica, introduciendo en el sistema procesal una variable de incertidumbre radical

que desafia los fundamentos mismos de la función probatoria⁵⁵⁷.

El fenómeno de los *deepfakes*, entendidos como contenidos audiovisuales generados o alterados mediante algoritmos de aprendizaje profundo que simulan con extraordinaria fidelidad la apariencia, voz y comportamiento de personas reales, no constituye simplemente un problema técnico de detección de falsificaciones. Representa, más bien, una transformación estructural en la economía epistémica de la prueba judicial, al socavar la distinción entre representación auténtica y simulación algorítmica con tal eficacia que ni la percepción humana ni los instrumentos periciales tradicionales resultan suficientes para garantizar la fiabilidad del material probatorio⁵⁵⁸. Esta crisis de autenticidad se proyecta simultáneamente sobre dos dimensiones procesales críticas: la admisibilidad de la evidencia y su valoración racional por parte del juzgador.

En el sistema del *common law*, particularmente en la tradición estadounidense, la autenticación de la prueba audiovisual descansa sobre las disposiciones de la Regla 901 de las *Federal Rules of Evidence*, cuya estructura normativa presupone que la identificación del origen y la verificación de la integridad del material pueden establecerse mediante testimonio directo o mediante la aplicación de teorías como la del testigo silencioso o la comunicación pictórica⁵⁵⁹. No obstante, la capacidad de los algoritmos generativos para producir grabaciones sintéticas indistinguibles de las auténticas invalida los supuestos sobre los cuales estas teorías fueron construidas, generando un vacío normativo que ni la jurisprudencia ni la doctrina han logrado colmar de manera satisfactoria⁵⁶⁰.

Paralelamente, en los sistemas de derecho continental, donde la valoración probatoria se rige por el principio de sana crítica judicial y por la

⁵⁵⁷ CHESNEY, R., & CITRON, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753-1820. <https://doi.org/10.15779/Z38RV0D15J>;

VACCARI, C., & CHADWICK, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1-13. <https://doi.org/10.1177/2056305120903408>

⁵⁵⁸ KIETZMANN, J., LEE, L. W., MCCARTHY, I. P., & KIETZMANN, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, 63(2), 135-146. <https://doi.org/10.1016/j.bushor.2019.11.006>; WISCHMEYER, T., & RADEMACHER, T. (Eds.). (2020). *Regulating artificial intelligence*.

Springer. <https://doi.org/10.1007/978-3-030-32361-5>

⁵⁵⁹ MCPEAK, A. (2021). Confronting the deepfake threat to trust in media, democratic deliberation, and evidence in court. *Vanderbilt Law Review*, 74(2), 445-504. <https://scholarship.law.vanderbilt.edu/vlr/vol74/iss2/4/>; PFEFFERKORN, R. (2020). Deepfakes in the courtroom. *Boston University Journal of Science & Technology Law*, 29(1), 245-276. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3617472

⁵⁶⁰ CITRON, D. K., & CHESNEY, R. (2018). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98(1), 147-155.

exigencia de motivación racional de las decisiones, el problema adquiere matices epistemológicos distintos, pero igualmente perturbadores. La sana crítica, entendida como el conjunto de reglas del correcto entendimiento humano que integran la lógica, la experiencia y los conocimientos científicos, presupone que el juzgador cuenta con instrumentos cognitivos adecuados para evaluar la fiabilidad de la evidencia presentada⁵⁶¹. Sin embargo, cuando la propia naturaleza del material probatorio se encuentra afectada por una manipulación algorítmica capaz de engañar tanto a la percepción sensorial como a los análisis forenses convencionales, la función valorativa del juez se ve radicalmente comprometida⁵⁶².

La literatura académica reciente ha comenzado a documentar las múltiples dimensiones de esta crisis procesal. Desde una perspectiva epistemológica, diversos autores han advertido que la proliferación de *deepfakes* genera un fenómeno de "dividendo del mentiroso", mediante el cual la mera posibilidad de que una

grabación haya sido manipulada permite a las partes procesales desacreditar evidencia auténtica, erosionando así la confianza general en la prueba audiovisual⁵⁶³. Desde un ángulo normativo, se ha señalado que los marcos regulatorios existentes, incluyendo las recientes iniciativas europeas como el Reglamento (UE) 2024/1689 sobre inteligencia artificial, establecen obligaciones de transparencia y etiquetado de contenidos sintéticos, pero carecen de mecanismos específicos de integración en los procedimientos judiciales de autenticación y valoración probatoria⁵⁶⁴.

La presente investigación tiene por objetivo analizar críticamente las deficiencias estructurales que la tecnología *deepfake* revela en los procesos de autenticación y valoración de la prueba audiovisual. En primer término, se propone realizar un análisis comparativo trans-sistémico de los marcos normativos y dogmáticos que rigen la autenticación y valoración de la prueba audiovisual en el *common law* y en el derecho continental, identificando tanto las convergencias

⁵⁶¹ TARUFFO, M. (2002). La prueba de los hechos. Trotta; TARUFFO, M. (2005). Conocimiento científico y estándares de prueba judicial. Boletín Mexicano de Derecho Comparado, 38(114), 1285-1312. <https://doi.org/10.22201/ijj.24484873e.2005.114.3770>

⁵⁶² FLORIDI, L. (2018). Artificial intelligence, deepfakes and a future of ectypes. Philosophy & Technology, 31(3), 317-321. <https://doi.org/10.1007/s13347-018-0325-3>; HILDEBRANDT, M. (2015). Smart technologies and the end(s) of law: Novel entanglements of law and technology. Edward Elgar Publishing.

⁵⁶³ CHESNEY, R., & CITRON, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753-1820. <https://doi.org/10.15779/Z38RV0D15J>.

⁵⁶⁴ VALE, M., & BORGES, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97-112. <https://doi.org/10.9785/crl-2021-220402>; WISCHMEYER, T., & RADEMACHER, T. (Eds.). (2020). Regulating artificial intelligence. Springer. <https://doi.org/10.1007/978-3-030-32361-5>.

como las divergencias estructurales entre ambas tradiciones jurídicas frente al desafío de los *deepfakes*. En segundo término, se busca formular un modelo híbrido de control judicial integral que articule los aportes técnicos del *common law* en materia de autenticación con la racionalidad epistémica del derecho continental en el ámbito de la valoración, proponiendo así una arquitectura procesal capaz de restablecer la legitimidad epistémica del sistema judicial en la era digital.

La relevancia de este estudio radica en la urgencia con que el derecho procesal contemporáneo debe responder a una transformación tecnológica que no solo afecta la fiabilidad de un medio probatorio específico, sino que amenaza con erosionar la confianza pública en la capacidad del sistema judicial para alcanzar la verdad procesal. Si la evidencia audiovisual, históricamente considerada uno de los medios más persuasivos y objetivos de prueba, puede ser fabricada o alterada con tal grado de sofisticación que resulta prácticamente indetectable, entonces la arquitectura epistémica sobre la cual descansa el proceso judicial requiere una reformulación profunda⁵⁶⁵.

Metodológicamente, esta investigación adopta un enfoque analítico-comparativo que integra el

análisis doctrinal de fuentes normativas primarias, la revisión sistemática de literatura académica especializada en derecho procesal, epistemología jurídica y estudios sobre inteligencia artificial, así como el examen de casos judiciales incipientes en los que se han planteado objeciones basadas en la posibilidad de manipulación mediante *deepfakes*. El marco teórico se construye a partir de las contribuciones seminales de la epistemología procesal contemporánea, particularmente los trabajos de Michele Taruffo sobre la prueba y la verdad en el proceso, así como las reflexiones de autores como Luigi Ferrajoli, Ronald Dworkin y Mireille Hildebrandt sobre la legitimidad epistémica y la función garantista del derecho en contextos de complejidad tecnológica.

La estructura del trabajo se organiza en cuatro secciones principales. La primera examina críticamente los marcos normativos vigentes de autenticación probatoria en el *common law* y el derecho continental, identificando sus fundamentos epistemológicos y sus limitaciones frente a la evidencia sintética. La segunda sección analiza los impactos cognitivos y epistemológicos que los *deepfakes* generan en la función judicial de búsqueda de la verdad, documentando fenómenos como el escepticismo

⁵⁶⁵ DELFINO, R. (2020). Deepfake porn and the impact to women and girls. University of San Francisco Law Research Paper. https://repository.uclawsf.edu/faculty_scholarship/761/; HARWELL, D. (2019). Scarlett Johansson on fake AI-generated sex videos:

'Nothing can stop someone from cutting and pasting my image'. The Washington Post. <https://www.washingtonpost.com/technology/2019/01/02/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/>

inducido y el dividendo del mentiroso. La tercera sección evalúa comparativamente las estrategias normativas emergentes en la Unión Europea y Estados Unidos, examinando su potencial de integración en los mecanismos procesales de control judicial. Finalmente, la cuarta sección formula un modelo híbrido de autenticación y valoración que aspira a reconciliar fiabilidad tecnológica y razonabilidad argumentativa en un marco procesal coherente y garantista.

A través de este análisis, se pretende demostrar que la crisis de la prueba audiovisual en la era de los *deepfakes* no puede resolverse mediante ajustes marginales a los marcos normativos existentes, sino que requiere una reformulación epistémica del rol del juez como garante de la verdad procesal, así como la construcción de nuevos estándares de autenticación y valoración que integren criterios de trazabilidad algorítmica, transparencia tecnológica y motivación judicial reforzada. Solo mediante esta reconceptualización procesal será posible preservar la legitimidad del sistema judicial frente a los desafíos que plantea la manipulación digital masiva de la evidencia audiovisual.

1. MARCOS NORMATIVOS Y DOGMÁTICOS DE AUTENTICACIÓN PROBATORIA: INSUFICIENCIAS ESTRUCTURALES FRENTE A LA EVIDENCIA SINTÉTICA

La arquitectura procesal contemporánea de autenticación probatoria se asienta sobre presupuestos epistemológicos que fueron elaborados en contextos tecnológicos radicalmente distintos a los actuales. Tanto en la tradición del *common law* como en el derecho continental, los mecanismos de verificación de autenticidad presuponen la existencia de una conexión estable entre el medio de prueba y el hecho que pretende representar, así como la posibilidad de que esta conexión pueda ser verificada mediante instrumentos cognitivos accesibles al juzgador o a los expertos técnicos. Sin embargo, la emergencia de tecnologías de síntesis audiovisual basadas en inteligencia artificial ha fracturado estos presupuestos, revelando la inadecuación estructural de los marcos normativos vigentes⁵⁶⁶.

En el sistema estadounidense, la Regla 901 de las Federal Rules of Evidence establece que la autenticación de la evidencia requiere la presentación de prueba suficiente para sostener que el material es lo que su proponente afirma que es. Esta

⁵⁶⁶ CHESNEY, R., & CITRON, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753-1820. <https://doi.org/10.15779/Z38RV0D15J>; MCPEAK, A. (2021). Confronting the deepfake

threat to trust in media, democratic deliberation, and evidence in court. *Vanderbilt Law Review*, 74(2), 445-504. <https://scholarship.law.vanderbilt.edu/vlr/vol74/iss2/4/>

disposición ha sido interpretada tradicionalmente bajo dos teorías procesales fundamentales. La teoría del testigo silencioso postula que ciertos dispositivos de grabación, como cámaras de seguridad o registros automáticos, pueden funcionar como testigos independientes cuya fiabilidad deriva de la correcta operación del mecanismo técnico. La teoría de la comunicación pictórica, por su parte, requiere que un testigo con conocimiento personal pueda atestiguar que la grabación representa fielmente los hechos que presenci⁵⁶⁷. Ambas teorías, sin embargo, parten del supuesto implícito de que las grabaciones audiovisuales constituyen registros de eventos reales y no simulaciones generadas algorítmicamente.

La sofisticación técnica de los *deepfakes* contemporáneos invalida estas premisas de manera radical. Las redes neuronales generativas adversarias son capaces de producir contenidos audiovisuales sintéticos que no solo replican con extraordinaria precisión las características físicas, vocales y gestuales de individuos

reales, sino que incorporan contextos ambientales, iluminación y artefactos visuales que imitan perfectamente las imperfecciones propias de las grabaciones auténticas⁵⁶⁸. Esta capacidad de mimetización tecnológica socava la teoría del testigo silencioso, puesto que el dispositivo de grabación ya no garantiza la autenticidad del contenido cuando este puede haber sido generado íntegramente mediante algoritmos. Simultáneamente, erosiona la teoría de la comunicación pictórica, dado que incluso un testigo presencial podría ser engañado por una simulación suficientemente convincente⁵⁶⁹.

La jurisprudencia estadounidense ha comenzado a enfrentar esta problemática de manera fragmentaria. En casos recientes, las partes procesales han planteado objeciones basadas en la posibilidad de que evidencia audiovisual pudiera haber sido manipulada mediante *deepfakes*, aunque los tribunales aún no han desarrollado una doctrina consistente para el tratamiento de estas objeciones⁵⁷⁰. La ausencia de criterios judiciales claros genera una

⁵⁶⁷ PFEFFERKORN, R. (2020). Deepfakes in the courtroom. *Boston University Journal of Science & Technology Law*, 29(1), 245-276. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3617472; GRIMM, P., Capra, D., & Joseph, G. (2018). *Evidence: A contemporary approach*. West Academic Publishing.

⁵⁶⁸ KIETZMANN, J., LEE, L. W., MCCARTHY, I. P., & KIETZMANN, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, 63(2), 135-146. <https://doi.org/10.1016/j.bushor.2019.11.006>; TOLOSANA, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face

manipulation and fake detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>.

⁵⁶⁹ CITRON, D. K., & CHESNEY, R. (2018). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98(1), 147-155.

⁵⁷⁰ DELFINO, R. (2020). Deepfake porn and the impact to women and girls. University of San Francisco Law Research Paper. https://repository.uclawsf.edu/faculty_scholarship/761/; HARWELL, D. (2019). Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and

situación de incertidumbre procesal en la que los jueces carecen de estándares operativos para evaluar la autenticidad de grabaciones potencialmente sintéticas, mientras que las partes enfrentan dificultades para cumplir con la carga probatoria de demostrar tanto la autenticidad como la falsificación de material audiovisual⁵⁷¹.

En el ámbito del derecho continental, el problema se manifiesta con características epistemológicas distintas pero igualmente perturbadoras. La tradición procesal civil, heredera del derecho romano-canónico y desarrollada particularmente en los sistemas jurídicos europeos y latinoamericanos, no establece requisitos formales de admisibilidad tan estrictos como los del *common law*, privilegiando en cambio un sistema de libre valoración de la prueba regido por el principio de sana crítica judicial. Este principio, consagrado en numerosos códigos procesales contemporáneos, exige que el juzgador valore la evidencia conforme a las reglas de la lógica, las máximas de la experiencia y los

conocimientos científicos, debiendo motivar racionalmente sus conclusiones.⁵⁷²

La sana crítica judicial presupone, sin embargo, que el juez dispone de instrumentos cognitivos adecuados para evaluar la fiabilidad del material probatorio. Cuando la evidencia audiovisual se encuentra potencialmente afectada por manipulación algorítmica capaz de engañar tanto a la percepción humana como a los análisis periciales convencionales, la función valorativa del juzgador se ve radicalmente comprometida⁵⁷³. El juez debe entonces enfrentar una paradoja epistémica: por un lado, no puede simplemente rechazar la evidencia audiovisual sin vulnerar el principio de amplitud probatoria y el derecho de las partes a presentar pruebas; por otro lado, tampoco puede valorarla sin disponer de criterios técnicos suficientes para determinar su autenticidad⁵⁷⁴.

Esta insuficiencia estructural se ve agravada por la complejidad técnica inherente a la detección de *deepfakes*. A diferencia de las manipulaciones

pasting my image'. The Washington Post. <https://www.washingtonpost.com/technology/2019/01/02/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/>

⁵⁷¹ MCPEAK, A. (2021). Confronting the deepfake threat to trust in media, democratic deliberation, and evidence in court. *Vanderbilt Law Review*, 74(2), 445-504. <https://scholarship.law.vanderbilt.edu/vlr/vol74/iss2/4/>

⁵⁷² TARUFFO, M. (2002). La prueba de los hechos. Trotta; TARUFFO, M. (2005). Conocimiento científico y estándares de

prueba judicial. *Boletín Mexicano de Derecho Comparado*, 38(114), 1285-1312. <https://doi.org/10.22201/ij.24484873e.2005.114.3770>.

⁵⁷³ FERRAJOLI, L. (2011). *Principia iuris: Teoría del derecho y de la democracia*. Trotta; GASCÓN ABELLÁN, M. (2004). *Los hechos en el derecho: Bases argumentales de la prueba*. Marcial Pons.

⁵⁷⁴ CARNELUTTI, F. (1950). *La prueba civil*. Ediciones Depalma; COUTURE, E. J. (1958). *Fundamentos del derecho procesal civil*. Ediciones Depalma.

fotográficas tradicionais, que dejaban huellas detectables mediante análisis forenses convencionales, los contenidos generados por redes neuronales profundas no presentan necesariamente artefactos o inconsistencias identificables mediante inspección visual o técnicas de análisis de imagen estándar⁵⁷⁵. Los métodos de detección más avanzados requieren el empleo de algoritmos de aprendizaje automático entrenados específicamente para identificar patrones sutiles en la estructura de los píxeles, los patrones de compresión o las inconsistencias en los metadatos, pero estos métodos enfrentan a su vez limitaciones significativas en términos de precisión, generalización y vulnerabilidad ante técnicas adversarias de evasión⁵⁷⁶.

Desde una perspectiva dogmática, la inadecuación de los marcos procesales tradicionales se manifiesta en la ausencia de criterios específicos para la autenticación de evidencia sintética. Tanto en el *common law* como en el derecho

continental, las categorías jurídicas de "autenticidad", "relevancia" y "valor probatorio" fueron elaboradas bajo el supuesto de que los medios de prueba constituyen representaciones de hechos reales, ya sea mediante testimonios humanos, documentos físicos o registros mecánicos. La posibilidad de que una grabación audiovisual pueda ser íntegramente fabricada mediante algoritmos, sin que exista ningún evento subyacente al cual la representación se refiera, cuestiona la aplicabilidad misma de estas categorías tradicionales⁵⁷⁷.

La literatura procesal contemporánea ha comenzado a identificar esta laguna normativa y dogmática. Diversos autores han señalado que los sistemas judiciales requieren el desarrollo de nuevos estándares de autenticación que integren criterios de trazabilidad digital, verificación de metadatos, análisis de cadena de custodia digital y empleo de técnicas forenses computacionales avanzadas⁵⁷⁸. Sin embargo, la mera incorporación de

⁵⁷⁵ TOLOSANA, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>; VERDOLIVA, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5), 910-932. <https://doi.org/10.1109/JSTSP.2020.3002101>.

⁵⁷⁶ NGUYEN, T. T., NGUYEN, C. M., NGUYEN, D. T., NGUYEN, D. T., & NAHAVANDI, S. (2019). Deep learning for deepfakes creation and detection: A survey. *arXiv preprint arXiv:1909.11573*. <https://doi.org/10.48550/arXiv.1909.11573>;

GÜERA, D., & DELP, E. J. (2018). Deepfake video detection using recurrent neural networks. *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, 1-6. <https://doi.org/10.1109/AVSS.2018.8639163>.

⁵⁷⁷ TARUFFO, M. (2005). Conocimiento científico y estándares de prueba judicial. *Boletín Mexicano de Derecho Comparado*, 38(114), 1285-1312. <https://doi.org/10.22201/ij.24484873e.2005.14.3770>; TWINING, W. (2006). *Rethinking evidence: Exploratory essays*. Cambridge University Press.

⁵⁷⁸ MCPEAK, A. (2021). *Confronting the deepfake threat to trust in media, democratic deliberation, and evidence in court*. Vanderbilt

herramientas técnicas resulta insuficiente si no se acompaña de una reformulación epistemológica del rol del juez como garante de la verdad procesal en contextos de alta complejidad tecnológica⁵⁷⁹.

La comparación trans-sistémica entre el *common law* y el derecho continental revela tanto convergencias como divergencias significativas en el tratamiento de esta problemática. Ambas tradiciones comparten la dificultad estructural de adaptar sus categorías procesales tradicionales a una forma de evidencia que desafía los presupuestos epistemológicos sobre los cuales fueron construidas. Sin embargo, divergen en sus mecanismos de respuesta institucional. El *common law*, con su énfasis en reglas de admisibilidad y estándares de autenticación, tiende a concentrar la problemática en la fase de admisión de la prueba, exigiendo que la parte proponente satisfaga requisitos específicos antes de que la evidencia pueda ser presentada al jurado. El derecho continental, con su principio de libre valoración y su énfasis en la motivación judicial, tiende a trasladar la problemática a la fase de valoración,

exigiendo que el juez argumente racionalmente las razones por las cuales considera que determinada evidencia es auténtica y fiable⁵⁸⁰.

Ninguno de estos enfoques resulta plenamente satisfactorio frente al desafío de los *deepfakes*. Un sistema que concentra el control en la fase de admisibilidad corre el riesgo de excluir evidencia potencialmente auténtica por temor a la falsificación, mientras que un sistema que delega excesivamente en la valoración judicial puede carecer de criterios técnicos suficientes para guiar la decisión del juzgador. La solución, según diversos autores, requiere una arquitectura procesal híbrida que integre mecanismos de control técnico en la fase de autenticación con criterios de valoración racional en la fase de apreciación probatoria⁵⁸¹.

Esta reformulación procesal debe considerar, además, las dimensiones sistémicas del problema. La proliferación de *deepfakes* no afecta únicamente casos aislados, sino que genera un riesgo estructural para la confianza pública en el sistema judicial. Cuando los ciudadanos perciben que las grabaciones audiovisuales pueden ser fabricadas

Law Review, 74(2), 445-504. <https://scholarship.law.vanderbilt.edu/vlr/vol74/iss2/4/>; PFEFFERKORN, R. (2020). Deepfakes in the courtroom. Boston University Journal of Science & Technology Law, 29(1), 245-276. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3617472.

⁵⁷⁹ HILDEBRANDT, M. (2015). Smart technologies and the end(s) of law: Novel entanglements of law and technology. Edward Elgar Publishing; FLORIDI, L. (2018). Artificial intelligence, deepfakes and a future of ectypes.

Philosophy & Technology, 31(3), 317-321. <https://doi.org/10.1007/s13347-018-0325-3>.

⁵⁸⁰ FERRER BELTRÁN, J. (2007). La valoración racional de la prueba. Marcial Pons; GASCÓN ABELLÁN, M. (2004). Los hechos en el derecho: Bases argumentales de la prueba. Marcial Pons.

⁵⁸¹ TARUFFO, M. (2009). La semplice verità: Il giudice e la costruzione dei fatti. Laterza; TWINING, W. (2006). Rethinking evidence: Exploratory essays. Cambridge University Press.

con facilidad y que los tribunales carecen de instrumentos adecuados para detectar estas falsificaciones, la legitimidad epistémica del proceso judicial se ve erosionada en su conjunto⁵⁸². Este fenómeno de desconfianza generalizada puede manifestarse tanto en la incredulidad frente a evidencia auténtica como en la credulidad frente a evidencia manipulada, socavando así la función epistémica fundamental del proceso judicial.

2. IMPACTOS COGNITIVOS Y EPISTEMOLÓGICOS EN LA FUNCIÓN JUDICIAL DE BÚSQUEDA DE LA VERDAD

La transformación tecnológica que representan los *deepfakes* no se limita a plantear desafíos técnicos de detección, sino que genera impactos profundos en las dimensiones cognitivas y epistemológicas de la función judicial. La capacidad de producir simulaciones audiovisuales indistinguibles de las grabaciones auténticas altera fundamentalmente la relación entre percepción sensorial,

creencia justificada y conocimiento procesal, socavando los mecanismos epistémicos tradicionales mediante los cuales jueces, peritos y testigos construyen sus convicciones sobre los hechos controvertidos⁵⁸³.

Desde una perspectiva epistemológica, el proceso judicial puede conceptualizarse como un sistema institucionalizado de producción de creencias justificadas sobre hechos pasados. Esta función epistémica se apoya en una compleja arquitectura de garantías procedimentales que incluyen el derecho a la contradicción, la inmediación en la apreciación de la prueba, la motivación de las decisiones y la posibilidad de control recursivo⁵⁸⁴. Sin embargo, todos estos mecanismos presuponen que los medios de prueba mantienen una relación de correspondencia verificable con los hechos que pretenden representar. Cuando esta correspondencia se ve comprometida por la capacidad de fabricación sintética, la arquitectura epistémica entera se desestabiliza⁵⁸⁵.

⁵⁸² DELFINO, R. (2020). Deepfake porn and the impact to women and girls. University of San Francisco Law Research Paper. https://repository.uclawsf.edu/faculty_scholarship/761/; VACCARI, C., & CHADWICK, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1-13. <https://doi.org/10.1177/2056305120903408>.

⁵⁸³ VACCARI, C., & CHADWICK, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news.

Social Media + Society, 6(1), 1-13. <https://doi.org/10.1177/2056305120903408>; KIETZMANN, J., LEE, L. W., MCCARTHY, I. P., & KIETZMANN, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, 63(2), 135-146. <https://doi.org/10.1016/j.bushor.2019.11.006>.

⁵⁸⁴ FERRAJOLI, L. (2011). Principia iuris: Teoría del derecho y de la democracia. Trotta; TARUFFO, M. (2002). La prueba de los hechos. Trotta.

⁵⁸⁵ FLORIDI, L. (2018). Artificial intelligence, deepfakes and a future of ectypes. *Philosophy & Technology*, 31(3), 317-321. <https://doi.org/10.1007/s13347-018-0325-3>;

El fenómeno psicológico de la credibilidad automática de la evidencia visual constituye uno de los fundamentos sobre los cuales descansa tradicionalmente la fuerza persuasiva de las grabaciones audiovisuales. Décadas de investigación en psicología cognitiva han demostrado que los seres humanos tienden a otorgar mayor credibilidad a la información presentada visualmente que a los testimonios verbales o la documentación escrita, efecto conocido como "sesgo de verosimilitud visual"⁵⁸⁶. Este sesgo cognitivo se ve dramáticamente explotado por los *deepfakes*, que aprovechan la tendencia humana a confiar en la evidencia visual para producir simulaciones convincentes que pueden engañar incluso a observadores entrenados⁵⁸⁷.

La literatura sobre psicología del testimonio ha documentado ampliamente las limitaciones de la percepción humana y la vulnerabilidad de la memoria frente a sugerencias y distorsiones. Sin embargo, el problema que plantean los *deepfakes* trasciende

estas limitaciones cognitivas tradicionales, puesto que no se trata simplemente de errores perceptivos o de memoria, sino de la fabricación deliberada de realidades alternativas mediante tecnología capaz de superar los mecanismos evolutivos de detección de engaño que los seres humanos han desarrollado a lo largo de milenios⁵⁸⁸.

El impacto cognitivo de los *deepfakes* se proyecta sobre múltiples actores del proceso judicial. Los jueces, encargados de valorar la evidencia y decidir sobre los hechos, enfrentan la dificultad de evaluar la autenticidad de material audiovisual sin disponer necesariamente de conocimientos técnicos especializados en análisis forense digital. Los peritos, cuya función consiste en proporcionar conocimientos especializados que excedan el dominio común del juzgador, se ven confrontados con la necesidad de emplear técnicas de detección cada vez más sofisticadas, cuya fiabilidad y margen de error no siempre pueden ser explicados de manera accesible al tribunal⁵⁸⁹. Los

HILDEBRANDT, M. (2015). *Smart technologies and the end(s) of law: Novel entanglements of law and technology*. Edward Elgar Publishing.

⁵⁸⁶ FARID, H. (2019). Creating, using, misusing, and detecting deep fakes. *Journal of Online Trust and Safety*, 1(4), 1-16. <https://doi.org/10.54501/jots.v1i4.56>; NIGHTINGALE, S. J., WADE, K. A., & WATSON, D. G. (2017). Can people identify original and manipulated photos of real-world scenes? *Cognitive Research: Principles and Implications*, 2(1), 1-21. <https://doi.org/10.1186/s41235-017-0067-2>.

⁵⁸⁷ VACCARI, C., & CHADWICK, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1-13. <https://doi.org/10.1177/2056305120903408>.

⁵⁸⁸ WESTERLUND, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11), 39-52. <https://doi.org/10.22215/timreview/1282>.

⁵⁸⁹ TOLOSANA, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information*

testigos, por su parte, pueden ser llamados a identificar si determinada grabación representa fielmente los hechos que presenciaron, pero su capacidad para detectar manipulaciones sutiles resulta extremadamente limitada⁵⁹⁰.

Uno de los fenómenos más preocupantes desde el punto de vista epistemológico es el denominado "dividendo del mentiroso", concepto propuesto por Robert Chesney y Danielle Citron para describir la situación en la cual la mera existencia de tecnología de manipulación audiovisual permite a las partes procesales desacreditar evidencia auténtica mediante la simple sugerencia de que podría tratarse de un *deepfake*⁵⁹¹. Este fenómeno genera una inversión perversa de la carga probatoria: mientras que tradicionalmente correspondía a la parte que impugnaba la autenticidad de un documento demostrar su falsificación, en el contexto de los

deepfakes basta con plantear la posibilidad de manipulación para sembrar duda razonable sobre material genuino⁵⁹².

El dividendo del mentiroso no constituye simplemente una estrategia procesal oportunista, sino que refleja una transformación estructural en la economía epistémica de la prueba judicial. En un entorno donde la tecnología de falsificación audiovisual se encuentra ampliamente disponible y su sofisticación aumenta constantemente, la presunción de autenticidad de la que históricamente gozaban las grabaciones visuales se ve erosionada de manera irreversible. Esta erosión genera dos riesgos procesales simétricos: por un lado, la posibilidad de que evidencia falsificada sea admitida y valorada como auténtica; por otro lado, la probabilidad de que evidencia genuina sea descartada por temor a la falsificación⁵⁹³.

Fusion, 64, 131-148.
<https://doi.org/10.1016/j.inffus.2020.06.014>;
VERDOLIVA, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5), 910-932.

<https://doi.org/10.1109/JSTSP.2020.3002101>.

⁵⁹⁰ NIGHTINGALE, S. J., WADE, K. A., & WATSON, D. G. (2017). Can people identify original and manipulated photos of real-world scenes? *Cognitive Research: Principles and Implications*, 2(1), 1-21.
<https://doi.org/10.1186/s41235-017-0067-2>;
FARID, H. (2019). Creating, using, misusing, and detecting deep fakes. *Journal of Online Trust and Safety*, 1(4), 1-16.
<https://doi.org/10.54501/jots.v1i4.56>.

⁵⁹¹ CHESNEY, R., & CITRON, D. K. (2019). Deep fakes: A looming challenge for privacy,

democracy, and national security. *California Law Review*, 107(6), 1753-1820.
<https://doi.org/10.15779/Z38RV0D15J>.

⁵⁹² CITRON, D. K., & CHESNEY, R. (2018). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98(1), 147-155.

⁵⁹³ DELFINO, R. (2020). Deepfake porn and the impact to women and girls. University of San Francisco Law Research Paper. https://repository.uclawsf.edu/faculty_scholarship/761/; HARWELL, D. (2019). Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image'. *The Washington Post*. <https://www.washingtonpost.com/technology/2019/01/02/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/>

La investigación empírica sobre la percepción pública de los *deepfakes* ha comenzado a documentar el fenómeno del "escepticismo inducido", mediante el cual el conocimiento de la existencia de tecnologías de manipulación audiovisual genera una desconfianza generalizada hacia toda evidencia visual, independientemente de su autenticidad real⁵⁹⁴. Este escepticismo no se limita a los actores judiciales profesionales, sino que se extiende a los potenciales jurados en sistemas que emplean juicio por jurados, así como a la opinión pública en general, con consecuencias significativas para la legitimidad social de las decisiones judiciales⁵⁹⁵.

Desde una perspectiva epistemológica procesal, estos impactos cognitivos plantean desafíos fundamentales para la teoría de la verdad en el proceso judicial. La concepción tradicional de la verdad procesal como correspondencia entre la reconstrucción fáctica realizada por el tribunal y los hechos realmente ocurridos presupone que existen

mecanismos epistémicos confiables para establecer dicha correspondencia. Sin embargo, cuando los medios de prueba pueden ser fabricados sintéticamente con tal grado de sofisticación que resultan indistinguibles de los registros auténticos, la noción misma de verdad procesal como correspondencia se ve radicalmente cuestionada⁵⁹⁶.

Algunos autores han propuesto abandonar la concepción correspondentista de la verdad procesal en favor de teorías coherentistas o pragmatistas, argumentando que en contextos de alta complejidad tecnológica el sistema judicial debe conformarse con establecer la coherencia interna del relato fáctico más que su correspondencia con una realidad externa verificable⁵⁹⁷. Sin embargo, esta propuesta enfrenta objeciones significativas desde el punto de vista de la legitimidad del proceso, puesto que renunciar a la aspiración de verdad como correspondencia podría socavar la función garantista del derecho

⁵⁹⁴ VACCARI, C., & CHADWICK, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1-13. <https://doi.org/10.1177/2056305120903408>; KIETZMANN, J., LEE, L. W., MCCARTHY, I. P., & KIETZMANN, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, 63(2), 135-146. <https://doi.org/10.1016/j.bushor.2019.11.006>.

⁵⁹⁵ DOBBER, T., METOUI, N., TRILLING, D., HELBERGER, N., & de VREESE, C. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *International Journal of Press/Politics*, 26(1), 69-91. <https://doi.org/10.1177/1940161220944364>; TERNOVSKI, J., Kalla, J., & Aronow, P. (2022).

The negative consequences of informing voters about deepfakes: Evidence from two survey experiments. *Journal of Online Trust and Safety*, 1(2), 1-21. <https://doi.org/10.54501/jots.v1i2.36>.

⁵⁹⁶ TARUFFO, M. (2005). Conocimiento científico y estándares de prueba judicial. *Boletín Mexicano de Derecho Comparado*, 38(114), 1285-1312. <https://doi.org/10.22201/ij.24484873e.2005.14.3770>; FERRER BELTRÁN, J. (2007). La valoración racional de la prueba. *Marcial Pons*.

⁵⁹⁷ HAACK, S. (2014). *Evidence matters: Science, proof, and truth in the law*. Cambridge University Press; GASCÓN ABELLÁN, M. (2004). *Los hechos en el derecho: Bases argumentales de la prueba*. Marcial Pons.

procesal y facilitar la impunidad o la condena injusta⁵⁹⁸.

La tensión entre verdad procesal y complejidad tecnológica se manifiesta también en la relación entre conocimiento experto y decisión judicial. Tradicionalmente, los sistemas procesales han resuelto la limitación cognitiva del juzgador mediante la incorporación de prueba pericial, confiando en que los expertos pueden proporcionar conocimientos especializados que permitan evaluar evidencia técnica compleja. Sin embargo, en el caso de los *deepfakes*, la propia fiabilidad de los métodos de detección se encuentra en constante evolución y debate, generando lo que algunos autores han denominado "incertidumbre epistémica de segundo orden": no solo existe incertidumbre sobre si determinada grabación es auténtica, sino también incertidumbre sobre la fiabilidad de los métodos empleados para determinar su autenticidad⁵⁹⁹.

Esta incertidumbre de segundo orden plantea desafíos particulares para la motivación judicial. Si el juez basa su decisión sobre la autenticidad de una grabación en un informe pericial cuya metodología y margen de error no comprende plenamente, la motivación

de su decisión corre el riesgo de convertirse en una mera delegación acrítica en el criterio del experto, vulnerando así el principio de responsabilidad judicial en la valoración de la prueba⁶⁰⁰. Por otra parte, si el juez pretende evaluar críticamente la metodología pericial sin disponer de conocimientos técnicos suficientes, su valoración puede resultar arbitraria o superficial, igualmente vulnerando la exigencia de racionalidad en la decisión probatoria⁶⁰¹.

La dimensión sistémica de estos impactos cognitivos y epistemológicos resulta particularmente preocupante cuando se considera la acumulación de efectos que la proliferación de *deepfakes* genera en la confianza pública hacia el sistema judicial. La legitimidad de las instituciones judiciales no descansa únicamente en la corrección técnica de sus decisiones, sino también en la percepción social de que el proceso judicial constituye un mecanismo confiable para la determinación de la verdad y la aplicación imparcial del derecho⁶⁰². Cuando la ciudadanía percibe que los tribunales carecen de instrumentos adecuados para

⁵⁹⁸ FERRAJOLI, L. (2011). *Principia iuris: Teoría del derecho y de la democracia*. Trotta; DWORKIN, R. (2011). *Justice for hedgehogs*. Harvard University Press.

⁵⁹⁹ HILDEBRANDT, M. (2015). *Smart technologies and the end(s) of law: Novel entanglements of law and technology*. Edward Elgar Publishing; JASANOFF, S. (1995). *Science at the bar: Law, science, and technology in America*. Harvard University Press.

⁶⁰⁰ TARUFFO, M. (2009). *La semplice verità: Il giudice e la costruzione dei fatti*. Laterza; FERRER BELTRÁN, J. (2007). *La valoración racional de la prueba*. Marcial Pons.

⁶⁰¹ GASCÓN ABELLÁN, M. (2004). *Los hechos en el derecho: Bases argumentales de la prueba*. Marcial Pons.

⁶⁰² TYLER, T. R. (2006). *Why people obey the law*. Princeton University Press; FRIEDMAN, L. M. (2016). *Impact: How law affects behavior*. Harvard University Press.

distinguir entre evidencia auténtica y falsificaciones sofisticadas, la confianza en la justicia se ve erosionada de manera estructural, con consecuencias que trascienden los casos individuales⁶⁰³.

La investigación empírica sobre legitimidad institucional ha demostrado que la percepción de competencia técnica del sistema judicial constituye uno de los factores determinantes de la confianza pública⁶⁰⁴. En este contexto, la incapacidad manifiesta de los tribunales para enfrentar el desafío de los *deepfakes* mediante marcos normativos y procedimientos adecuados puede generar un círculo vicioso de deslegitimación: la ausencia de respuestas institucionales efectivas alimenta el escepticismo público, el cual a su vez refuerza el dividendo del mentiroso y facilita el empleo estratégico de objeciones basadas en la posibilidad de manipulación⁶⁰⁵.

Desde una perspectiva comparada, resulta relevante observar que los impactos cognitivos y

epistemológicos de los *deepfakes* se manifiestan de manera diferencial en sistemas con juicio por jurados y en sistemas con decisión judicial profesional. En los sistemas anglosajones que emplean jurados populares, la vulnerabilidad cognitiva frente a la manipulación audiovisual se ve agravada por la falta de entrenamiento especializado de los jurados y su mayor susceptibilidad a sesgos perceptivos y emocionales⁶⁰⁶. En los sistemas continentales con jueces profesionales, aunque la formación jurídica especializada puede proporcionar mayor capacidad crítica, persiste la limitación fundamental de que los jueces carecen típicamente de conocimientos técnicos avanzados en análisis forense digital y ciencias de la computación⁶⁰⁷.

Esta divergencia entre sistemas procesales sugiere que las estrategias de respuesta institucional deben adaptarse a las características específicas de cada tradición jurídica. Los sistemas con jurados pueden requerir desarrollos más robustos en

⁶⁰³ DELFINO, R. (2020). Deepfake porn and the impact to women and girls. University of San Francisco Law Research Paper. https://repository.uclawsf.edu/faculty_scholarship/761/; VACCARI, C., & CHADWICK, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1-13. <https://doi.org/10.1177/2056305120903408>.

⁶⁰⁴ TYLER, T. R., & HUO, Y. J. (2002). Trust in the law: Encouraging public cooperation with the police and courts. Russell Sage Foundation.

⁶⁰⁵ CITRON, D. K., & CHESNEY, R. (2018). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign*

Affairs, 98(1), 147-155; CHESNEY, R., & CITRON, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753-1820. <https://doi.org/10.15779/Z38RV0D15J>.

⁶⁰⁶ DELFINO, R. (2020). Deepfake porn and the impact to women and girls. University of San Francisco Law Research Paper. https://repository.uclawsf.edu/faculty_scholarship/761/; FARID, H. (2019). Creating, using, misusing, and detecting deep fakes. *Journal of Online Trust and Safety*, 1(4), 1-16. <https://doi.org/10.54501/jots.v1i4.56>.

⁶⁰⁷ TARUFFO, M. (2009). La semplice verità: Il giudice e la costruzione dei fatti. Laterza.

las reglas de admisibilidad y en las instrucciones que los jueces proporcionan a los jurados sobre la evaluación de evidencia audiovisual potencialmente manipulada. Los sistemas con jueces profesionales pueden necesitar mayor énfasis en la formación judicial continua, el desarrollo de protocolos de valoración pericial y el establecimiento de estándares de motivación reforzada cuando se trata de evidencia digital compleja⁶⁰⁸.

La dimensión temporal de la problemática añade complejidad adicional al análisis epistemológico. A diferencia de otros desafíos probatorios que emergen gradualmente y permiten la adaptación progresiva de las instituciones judiciales, la tecnología de *deepfakes* ha experimentado una aceleración exponencial en su sofisticación y accesibilidad en un período extraordinariamente breve⁶⁰⁹. Esta velocidad de transformación tecnológica genera un desfase estructural entre el ritmo de evolución

del derecho procesal, que típicamente opera mediante adaptaciones incrementales y consolidación jurisprudencial paulatina, y la rapidez con que nuevas técnicas de manipulación se vuelven disponibles y difundidas⁶¹⁰.

Este desfase temporal plantea interrogantes fundamentales sobre la capacidad del derecho procesal para cumplir su función epistémica en contextos de cambio tecnológico acelerado. Si las categorías procesales y los mecanismos de control judicial requieren décadas para consolidarse mediante la práctica jurisprudencial y la reflexión doctrinal, mientras que las tecnologías de manipulación evolucionan en escalas temporales de meses o años, el sistema judicial corre el riesgo de encontrarse en un estado de obsolescencia permanente, siempre intentando adaptarse a tecnologías que ya han sido superadas por desarrollos más recientes⁶¹¹.

⁶⁰⁸ MCPEAK, A. (2021). Confronting the deepfake threat to trust in media, democratic deliberation, and evidence in court. *Vanderbilt Law Review*, 74(2), 445-504. <https://scholarship.law.vanderbilt.edu/vlr/vol74/iss2/4/>; PFEFFERKORN, R. (2020). Deepfakes in the courtroom. *Boston University Journal of Science & Technology Law*, 29(1), 245-276. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3617472.

⁶⁰⁹ TOLOSANA, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>; VERDOLIVA, L. (2020). Media forensics and

deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5), 910-932.

<https://doi.org/10.1109/JSTSP.2020.3002101>.

⁶¹⁰ HILDEBRANDT, M. (2015). *Smart technologies and the end(s) of law: Novel entanglements of law and technology*. Edward Elgar Publishing.

⁶¹¹ FLORIDI, L. (2018). Artificial intelligence, deepfakes and a future of ectypes. *Philosophy & Technology*, 31(3), 317-321. <https://doi.org/10.1007/s13347-018-0325-3>; WISCHMEYER, T., & RADEMACHER, T. (Eds.). (2020). *Regulating artificial intelligence*. Springer. <https://doi.org/10.1007/978-3-030-32361-5>;

3. ESTRATEGIAS NORMATIVAS EMERGENTES: INTEGRACIÓN DE LA REGULACIÓN DE INTELIGENCIA ARTIFICIAL EN LOS MECANISMOS PROCESALES

La conciencia creciente sobre los riesgos que plantean los *deepfakes* y otras formas de contenido sintético ha generado respuestas regulatorias en diversas jurisdicciones, particularmente en la Unión Europea y Estados Unidos. Sin embargo, estas iniciativas normativas han tendido a enfocarse en aspectos relacionados con la transparencia, el etiquetado y la responsabilidad de los proveedores de servicios digitales, sin desarrollar plenamente los mecanismos específicos de integración en los procedimientos judiciales de autenticación y valoración probatoria⁶¹².

El Reglamento (UE) 2024/1689 sobre inteligencia artificial, conocido como AI Act, constituye el marco regulatorio más ambicioso y comprehensivo desarrollado hasta la fecha para el gobierno de sistemas de inteligencia artificial. Este instrumento normativo establece un enfoque basado en riesgos, clasificando los sistemas de IA según su potencial de

daño y estableciendo obligaciones diferenciadas según el nivel de riesgo identificado⁶¹³. El artículo 3(60) del Reglamento define específicamente los *deepfakes* como contenido de imagen, audio o video generado o manipulado por IA que se asemeja apreciablemente a personas, objetos, lugares o entidades existentes, y que parecería falsamente auténtico o veraz a una persona.

Las obligaciones de transparencia establecidas por el AI Act resultan particularmente relevantes para el contexto procesal. El Reglamento impone a los proveedores de sistemas generativos la obligación de diseñar sus modelos de manera que el contenido sintético generado sea identificable como tal, mediante marcas de agua técnicas u otros mecanismos de identificación. Asimismo, los usuarios de sistemas generativos que produzcan contenido audiovisual deben divulgar que dicho contenido ha sido generado o manipulado artificialmente⁶¹⁴. Estas obligaciones podrían, en principio, facilitar significativamente la tarea de autenticación judicial, al establecer presunciones de manipulación cuando el contenido no incluye las marcas requeridas o cuando se verifica la

⁶¹² VEALE, M., & BORGESIU, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97-112. <https://doi.org/10.9785/cr-2021-220402>; HILDEBRANDT, M. (2020). *Law for computer scientists and other folk*. Oxford University Press.

⁶¹³ VEALE, M., & BORGESIU, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97-112. <https://doi.org/10.9785/cr-2021-220402>;

FLORIDI, L. (2021). The European legislation on AI: A brief analysis of its philosophical approach. *Philosophy & Technology*, 34(4), 215-222. <https://doi.org/10.1007/s13347-021-00460-9>.

⁶¹⁴ WISCHMEYER, T., & RADEMACHER, T. (Eds.). (2020). *Regulating artificial intelligence*. Springer. <https://doi.org/10.1007/978-3-030-32361-5>; BRADFORD, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.

presencia de indicadores técnicos de generación sintética.

Sin embargo, la efectividad de estas disposiciones para los propósitos procesales enfrenta limitaciones significativas. En primer lugar, las obligaciones del AI Act se aplican a los proveedores y usuarios legítimos de sistemas de IA, pero no afectan necesariamente a actores maliciosos que empleen tecnología de generación sintética con propósitos de fraude procesal. Los *deepfakes* utilizados deliberadamente para fabricar evidencia falsa no incluirán voluntariamente las marcas de agua o indicadores de transparencia requeridos por la regulación⁶¹⁵. En segundo lugar, la implementación técnica de mecanismos de marcado y detección enfrenta desafíos tecnológicos considerables, puesto que las marcas de agua digitales pueden ser removidas o alteradas mediante técnicas adversarias, y los indicadores de contenido sintético pueden resultar insuficientes frente a técnicas de generación cada vez más sofisticadas⁶¹⁶.

En el contexto estadounidense, la respuesta regulatoria ha sido más fragmentada, con iniciativas legislativas a nivel federal y estatal que

abordan aspectos específicos de los *deepfakes* sin configurar un marco comprehensivo comparable al AI Act europeo. Diversas propuestas legislativas federales, como la *Deepfake Report Act* y la *Deepfake Accountability Act*, han buscado establecer obligaciones de divulgación y mecanismos de responsabilidad civil y penal para la creación y distribución de *deepfakes* maliciosos⁶¹⁷. Sin embargo, ninguna de estas propuestas ha logrado hasta el momento establecer modificaciones específicas a las Federal Rules of Evidence o desarrollar estándares procesales claros para la autenticación de evidencia potencialmente manipulada.

Algunos estados estadounidenses han adoptado legislación específica sobre *deepfakes*, típicamente enfocada en contextos de pornografía no consensual, fraude electoral o suplantación de identidad. California, Texas y Virginia, entre otros estados, han promulgado leyes que criminalizan la creación y distribución de ciertos tipos de *deepfakes* o establecen causas de acción civil para las víctimas de manipulación audiovisual no

⁶¹⁵ VEALE, M., & BORGESIU, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97-112. <https://doi.org/10.9785/cr-2021-220402>.

⁶¹⁶ TOLOSANA, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>.

⁶¹⁷ CHESNEY, R., & CITRON, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753-1820. <https://doi.org/10.15779/Z38RV0D15J>; MCPEAK, A. (2021). Confronting the deepfake threat to trust in media, democratic deliberation, and evidence in court. *Vanderbilt Law Review*, 74(2), 445-504. <https://scholarship.law.vanderbilt.edu/vlr/vol74/iss2/4/>.

autorizada⁶¹⁸. Sin embargo, estas legislaciones estatales tampoco han desarrollado sistemáticamente los mecanismos procesales específicos para la autenticación y valoración de evidencia audiovisual en procedimientos judiciales, limitándose generalmente a establecer prohibiciones sustantivas sin abordar las cuestiones probatorias asociadas.

La comparación entre el enfoque europeo y el estadounidense revela diferencias significativas en la filosofía regulatoria. El modelo europeo, ejemplificado por el AI Act, adopta una aproximación *ex ante* basada en obligaciones de transparencia, trazabilidad y supervisión regulatoria de los sistemas de IA antes de su despliegue en el mercado. Este enfoque busca prevenir los daños potenciales mediante el establecimiento de estándares técnicos y organizacionales que los proveedores deben cumplir⁶¹⁹. El modelo estadounidense, por el contrario, tiende hacia una aproximación *ex post* basada en responsabilidad civil y penal después de que el daño se ha producido, confiando más en la litigación y en los

mecanismos de mercado para incentivar el comportamiento responsable⁶²⁰.

Ambos enfoques presentan ventajas y limitaciones desde la perspectiva del derecho procesal. El modelo europeo de transparencia *ex ante* podría facilitar la tarea de autenticación judicial al establecer presunciones basadas en la presencia o ausencia de marcas técnicas requeridas, pero enfrenta desafíos de implementación técnica y de aplicación frente a actores maliciosos. El modelo estadounidense de responsabilidad *ex post* proporciona incentivos para evitar la creación de *deepfakes* dañinos, pero no resuelve directamente la problemática probatoria de cómo los tribunales deben evaluar la autenticidad de evidencia potencialmente manipulada cuando esta es presentada en el proceso⁶²¹.

La integración de estas estrategias regulatorias en los mecanismos procesales específicos requiere desarrollos adicionales que hasta el momento no se han producido de manera sistemática. Desde una perspectiva procesal, resulta

⁶¹⁸ CITRON, D. K., & CHESNEY, R. (2018). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98(1), 147-155; AJDER, H., PATRINI, G., CAVALLI, F., & CULLEN, L. (2019). The state of deepfakes: Landscape, threats, and impact. *Deepttrace*.

https://regmedia.co.uk/2019/10/08/deepfake_report.pdf.

⁶¹⁹ VEALE, M., & BORGESIU, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97-112. <https://doi.org/10.9785/cr-2021->

220402; BRADFORD, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.

⁶²⁰ CITRON, D. K., & CHESNEY, R. (2018). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98(1), 147-155.

⁶²¹ MCPEAK, A. (2021). Confronting the deepfake threat to trust in media, democratic deliberation, and evidence in court. *Vanderbilt Law Review*, 74(2), 445-504. <https://scholarship.law.vanderbilt.edu/vlr/vol74/iss2/4/>.

necesario traducir las obligaciones de transparencia establecidas por regulaciones como el AI Act en estándares operativos de autenticación que los tribunales puedan aplicar. Esto implica establecer presunciones probatorias, cargas de la prueba y estándares de persuasión específicos para evidencia audiovisual digital⁶²².

Una posible aproximación consiste en establecer un régimen de presunción de autenticidad refutable para contenido audiovisual que incluya los indicadores de transparencia requeridos por la regulación, junto con una presunción de manipulación también refutable para contenido que carezca de tales indicadores. Este régimen de presunciones invertiría estratégicamente la carga probatoria según el cumplimiento de las obligaciones regulatorias, incentivando así tanto la adopción de tecnologías de marcado como la producción de evidencia sobre la autenticidad del material presentado⁶²³. Sin embargo, la implementación de este régimen requeriría modificaciones legislativas específicas a los códigos procesales o reglas de evidencia, las cuales hasta el momento no se han producido en la mayoría de las jurisdicciones.

Otra dimensión relevante de las estrategias normativas emergentes se relaciona con el establecimiento de obligaciones de conservación y trazabilidad de metadatos asociados al contenido audiovisual. El AI Act establece requisitos de documentación para sistemas de alto riesgo, incluyendo el mantenimiento de registros sobre los datos de entrenamiento, la arquitectura del modelo y los parámetros operacionales⁶²⁴. Desde una perspectiva procesal, estos requisitos de documentación podrían proporcionar elementos valiosos para la autenticación, al permitir la verificación de la cadena de custodia digital y la trazabilidad del proceso de generación o captura del contenido audiovisual.

Sin embargo, la traducción de estas obligaciones regulatorias en mecanismos procesales efectivos enfrenta desafíos prácticos considerables. Los metadatos pueden ser alterados o falsificados, y su preservación depende de la cooperación de múltiples actores en la cadena de distribución digital⁶²⁵. Además, en contextos adversariales donde una parte tiene incentivos para fabricar evidencia falsa, no puede esperarse que voluntariamente

⁶²² FERRER BELTRÁN, J. (2007). La valoración racional de la prueba. Marcial Pons; GASCÓN ABELLÁN, M. (2004). Los hechos en el derecho: Bases argumentales de la prueba. Marcial Pons.

⁶²³ TARUFFO, M. (2009). La semplice verità: Il giudice e la costruzione dei fatti. Laterza.

⁶²⁴ WISCHMEYER, T., & RADEMACHER, T. (Eds.). (2020). Regulating artificial intelligence.

Springer. <https://doi.org/10.1007/978-3-030-32361-5>.

⁶²⁵ PFEFFERKORN, R. (2020). Deepfakes in the courtroom. Boston University Journal of Science & Technology Law, 29(1), 245-276. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3617472.

preserve metadatos que revelen la manipulación. Por tanto, los mecanismos de autenticación procesal no pueden depender exclusivamente de la disponibilidad de metadatos, sino que deben incorporar también técnicas forenses de análisis del contenido mismo⁶²⁶.

La convergencia regulatoria entre jurisdicciones constituye otro aspecto relevante de las estrategias normativas emergentes. El AI Act europeo ha sido diseñado con aspiraciones de efecto extraterritorial, aplicándose no solo a proveedores establecidos en la Unión Europea sino también a aquellos que ofrecen servicios en el mercado europeo⁶²⁷. Este enfoque regulatorio podría generar efectos de armonización global, incentivando a proveedores de otras jurisdicciones a adoptar estándares compatibles con los europeos para facilitar el acceso al mercado común. Sin embargo, la efectividad de esta convergencia dependerá de la capacidad de aplicación extraterritorial y de la voluntad de otras jurisdicciones para adoptar enfoques regulatorios similares⁶²⁸.

Desde una perspectiva procesal comparada, la armonización

regulatoria internacional en materia de inteligencia artificial podría facilitar significativamente el tratamiento de evidencia audiovisual en casos transnacionales. La cooperación judicial internacional y el reconocimiento mutuo de decisiones judiciales dependen en parte de la confianza en que las jurisdicciones involucradas aplican estándares procedimentales equivalentes. Si diferentes jurisdicciones adoptan enfoques radicalmente divergentes para la autenticación de evidencia digital, la cooperación internacional en materia probatoria se verá dificultada⁶²⁹.

4. HACIA UN MODELO HÍBRIDO DE CONTROL JUDICIAL INTEGRAL: AUTENTICACIÓN TÉCNICO-JURÍDICA Y VALORACIÓN RACIONAL AMPLIADA

La insuficiencia de los marcos normativos vigentes y la inadecuación de las estrategias regulatorias emergentes para abordar comprehensivamente la problemática procesal de los *deepfakes* revelan la necesidad de desarrollar un modelo procesal innovador que integre los aportes de ambas tradiciones jurídicas

⁶²⁶ TOLOSANA, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>.

⁶²⁷ BRADFORD, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.

⁶²⁸ VEALE, M., & BORGESIU, F. Z. (2021). *Demystifying the Draft EU Artificial Intelligence*

Act. Computer Law Review International, 22(4), 97-112. <https://doi.org/10.9785/cr-2021-220402>.

⁶²⁹ TARUFFO, M. (2005). Conocimiento científico y estándares de prueba judicial. *Boletín Mexicano de Derecho Comparado*, 38(114), 1285-1312. <https://doi.org/10.22201/ij.24484873e.2005.14.3770>.

y responda simultáneamente a las dimensiones técnicas y epistemológicas del desafío. El modelo que se propone a continuación articula dos fases complementarias: una etapa de control técnico-jurídico de autenticidad basada en criterios verificables de trazabilidad, y una fase subsecuente de valoración racional fundamentada en una concepción ampliada de la sana crítica judicial⁶³⁰.

El modelo propuesto parte de la premisa epistemológica de que la autenticación y la valoración constituyen funciones procesales diferenciadas aunque interrelacionadas. La autenticación responde a la pregunta de si la evidencia es lo que su proponente afirma que es, estableciendo así la admisibilidad del material probatorio. La valoración, por su parte, determina qué peso o fuerza probatoria debe asignarse a la evidencia autenticada para la construcción de la convicción judicial sobre los hechos⁶³¹. Esta distinción conceptual resulta particularmente relevante en el contexto de los *deepfakes*, puesto que el material audiovisual puede ser auténtico en el sentido de que efectivamente fue capturado por determinado dispositivo en

determinadas circunstancias, pero manipulado en aspectos específicos que afectan su fiabilidad como representación de los hechos controvertidos.

La primera fase del modelo, denominada control técnico-jurídico de autenticidad, establece un conjunto estructurado de criterios que deben ser verificados antes de que la evidencia audiovisual pueda ser admitida en el proceso. Estos criterios se organizan en tres niveles progresivos de exigencia, adaptados a las características específicas del caso y al nivel de controversia sobre la autenticidad del material. El primer nivel, aplicable a casos donde no existe impugnación específica de autenticidad, requiere la verificación de elementos básicos: identificación de la fuente del material, fecha y circunstancias de captura, y cadena de custodia desde la captura hasta la presentación judicial⁶³².

El segundo nivel, aplicable cuando existe impugnación genérica de autenticidad o cuando el material presenta características que suscitan dudas razonables, exige la presentación de análisis técnicos sobre la integridad del archivo digital. Esto incluye la verificación de

⁶³⁰ TARUFFO, M. (2009). *La semplice verità: Il giudice e la costruzione dei fatti*. Laterza; FERRER BELTRÁN, J. (2007). *La valoración racional de la prueba*. Marcial Pons.

⁶³¹ GASCÓN ABELLÁN, M. (2004). *Los hechos en el derecho: Bases argumentales de la prueba*. Marcial Pons; FERRAJOLI, L. (2011). *Principia iuris: Teoría del derecho y de la democracia*. Trotta.

⁶³² MCPEAK, A. (2021). *Confronting the deepfake threat to trust in media, democratic deliberation, and evidence in court*. *Vanderbilt Law Review*, 74(2), 445-504. <https://scholarship.law.vanderbilt.edu/vlr/vol74/iss2/4/>; PFEFFERKORN, R. (2020). *Deepfakes in the courtroom*. *Boston University Journal of Science & Technology Law*, 29(1), 245-276. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3617472.

metadatos, el análisis de la historia de modificaciones del archivo, la detección de inconsistencias en patrones de compresión, y el examen de posibles artefactos o anomalías visuales o auditivas⁶³³. Este nivel requiere típicamente la intervención de peritos especializados en análisis forense digital, cuyos informes deben especificar la metodología empleada, las limitaciones técnicas del análisis y el margen de error de las conclusiones.

El tercer nivel, aplicable en casos donde existe sospecha fundada de manipulación mediante *deepfakes* o donde la evidencia audiovisual constituye elemento probatorio crucial para la decisión, demanda la aplicación de técnicas avanzadas de detección algorítmica. Esto incluye el empleo de redes neuronales especializadas en la identificación de contenido sintético, el análisis de patrones biométricos y microexpresiones, la verificación de coherencia física en movimientos y reflejos, y la evaluación de consistencia en características ambientales⁶³⁴. Este nivel máximo de

escrutinio reconoce que la detección de *deepfakes* sofisticados requiere instrumentos tecnológicos comparables en sofisticación a los empleados para su creación.

La estructura de niveles progresivos responde a la necesidad de equilibrar rigor epistémico y eficiencia procesal. Exigir análisis forenses exhaustivos para toda evidencia audiovisual presentada en cualquier procedimiento generaría costos prohibitivos y dilaciones injustificadas en casos donde la autenticidad no se encuentra genuinamente controvertida. Por el contrario, aplicar únicamente verificaciones superficiales en todos los casos expondría al proceso judicial a riesgos significativos de manipulación en aquellas situaciones donde las partes tienen incentivos y capacidad para fabricar evidencia falsa⁶³⁵.

La asignación de la carga probatoria en cada nivel constituye un aspecto crucial del modelo propuesto. En el primer nivel, corresponde a la parte que presenta la evidencia

⁶³³ TOLOSANA, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>; VERDOLIVA, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5), 910-932. <https://doi.org/10.1109/JSTSP.2020.3002101>.

⁶³⁴ NGUYEN, T. T., NGUYEN, C. M., NGUYEN, D. T., NGUYEN, D. T., & NAHAVANDI, S. (2019). Deep learning for deepfakes creation and detection: A survey. *arXiv preprint*

arXiv:1909.11573.

<https://doi.org/10.48550/arXiv.1909.11573>; GÜERA, D., & DELP, E. J. (2018). Deepfake video detection using recurrent neural networks. *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, 1-6. <https://doi.org/10.1109/AVSS.2018.8639163>.

⁶³⁵ TARUFFO, M. (2005). Conocimiento científico y estándares de prueba judicial. *Boletín Mexicano de Derecho Comparado*, 38(114), 1285-1312. <https://doi.org/10.22201/ij.24484873e.2005.114.3770>; FERRER BELTRÁN, J. (2007). La valoración racional de la prueba. *Marcial Pons*.

proporcionar la información básica sobre origen y cadena de custodia, tarea que no difiere sustancialmente de las obligaciones tradicionales en materia de autenticación de documentos. En el segundo nivel, cuando existe impugnación, la carga se comparte: la parte impugnante debe especificar los motivos concretos de duda sobre la autenticidad, mientras que la parte proponente debe proporcionar elementos técnicos que disipen esas dudas. En el tercer nivel, ante sospecha fundada de manipulación, el tribunal puede ordenar de oficio la realización de análisis periciales especializados, reconociendo así que la determinación de autenticidad en casos complejos no puede depender exclusivamente de las capacidades económicas y técnicas de las partes⁶³⁶.

La segunda fase del modelo, denominada valoración racional ampliada, opera una vez que la evidencia ha superado los controles de autenticación y ha sido admitida formalmente en el proceso. Esta fase reconoce que incluso material audiovisual auténtico puede presentar diferentes grados de fiabilidad según múltiples factores: las condiciones de captura, la calidad técnica del registro, la posibilidad de manipulación parcial que no fue detectada en la fase de autenticación, y la coherencia del

contenido con otras evidencias disponibles⁶³⁷.

La valoración racional ampliada incorpora los principios tradicionales de la sana crítica judicial, pero los reformula para contextos de complejidad tecnológica. El criterio de lógica exige que la inferencia desde la evidencia audiovisual hacia las conclusiones fácticas sea deductiva o inductivamente válida, evitando falacias argumentativas. Sin embargo, en el contexto de los *deepfakes*, este criterio debe incluir también la evaluación de la coherencia técnica del material: si determinada grabación presenta características físicas o tecnológicas inconsistentes con las circunstancias alegadas de su captura, la inferencia basada en dicha grabación se ve debilitada independientemente de que el material haya superado los controles de autenticación⁶³⁸.

El criterio de experiencia común requiere que la valoración se fundamente en regularidades empíricas conocidas sobre el comportamiento humano, los fenómenos físicos y los procesos técnicos. En el contexto de evidencia potencialmente manipulada, este criterio debe incorporar el conocimiento sobre las capacidades actuales de las tecnologías de síntesis audiovisual, las técnicas habituales de

⁶³⁶ GASCÓN ABELLÁN, M. (2004). Los hechos en el derecho: Bases argumentales de la prueba. Marcial Pons.

⁶³⁷ TARUFFO, M. (2009). La semplice verità: Il giudice e la costruzione dei fatti. Laterza; FERRER BELTRÁN, J. (2007). La valoración racional de la prueba. Marcial Pons.

⁶³⁸ HILDEBRANDT, M. (2015). Smart technologies and the end(s) of law: Novel entanglements of law and technology. Edward Elgar Publishing; FLORIDI, L. (2018). Artificial intelligence, deepfakes and a future of ectypes. *Philosophy & Technology*, 31(3), 317-321. <https://doi.org/10.1007/s13347-018-0325-3>.

manipulación y las limitaciones de los métodos de detección. El juez debe poder considerar, como parte de las máximas de experiencia relevantes, que existen tecnologías capaces de fabricar grabaciones convincentes, que ciertos actores tienen incentivos para emplear estas tecnologías en contextos procesales, y que la ausencia de indicadores detectables de manipulación no garantiza necesariamente la autenticidad del material⁶³⁹.

El criterio de conocimiento científico adquiere particular relevancia en la valoración de evidencia digital compleja. El juez debe evaluar críticamente los informes periciales presentados, considerando la solidez de la metodología empleada, el consenso científico sobre la fiabilidad de las técnicas aplicadas, y las limitaciones explícitamente reconocidas por los peritos. Esta evaluación crítica no implica que el juez deba poseer conocimientos técnicos equivalentes a los del perito, pero sí requiere que pueda identificar inconsistencias lógicas, evaluar la razonabilidad de las conclusiones en relación con las premisas, y detectar posibles sesgos o conflictos de interés⁶⁴⁰.

La motivación de la decisión probatoria constituye el mecanismo

fundamental mediante el cual el juez hace explícito el proceso racional de valoración. En casos que involucran evidencia audiovisual potencialmente manipulada, la exigencia de motivación debe ser reforzada, requiriendo que el juez explicita: los elementos técnicos considerados para evaluar la autenticidad; las razones por las cuales consideró suficientes o insuficientes los análisis periciales presentados; la evaluación de coherencia entre la evidencia audiovisual y otros elementos probatorios disponibles; y las inferencias específicas que fundamentan sus conclusiones fácticas⁶⁴¹.

Este estándar de motivación reforzada cumple múltiples funciones procesales. En primer lugar, facilita el control recursivo de la decisión, permitiendo que tribunales superiores evalúen la racionalidad del proceso valorativo. En segundo lugar, genera incentivos para que el juez de primera instancia realice una valoración cuidadosa y reflexiva. En tercer lugar, aumenta la transparencia del proceso judicial y facilita la aceptación social de las decisiones, al hacer explícitos

⁶³⁹ VACCARI, C., & CHADWICK, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1-13. <https://doi.org/10.1177/2056305120903408>.

⁶⁴⁰ TARUFFO, M. (2002). La prueba de los hechos. Trotta; JASANOFF, S. (1995). Science at

the bar: Law, science, and technology in America. Harvard University Press.

⁶⁴¹ FERRER BELTRÁN, J. (2007). La valoración racional de la prueba. Marcial Pons; GASCÓN ABELLÁN, M. (2004). Los hechos en el derecho: Bases argumentales de la prueba. Marcial Pons.

los criterios empleados para evaluar evidencia compleja⁶⁴².

El modelo propuesto incorpora también mecanismos de actualización continua para responder a la evolución tecnológica. Reconociendo que las técnicas de generación y detección de *deepfakes* se encuentran en constante desarrollo, se establece que los criterios técnicos de autenticación y los estándares de valoración deben ser revisados periódicamente por comisiones especializadas que integren expertos en derecho procesal, ciencias de la computación y epistemología jurídica. Esta institucionalización de la actualización normativa busca reducir el desfase temporal entre la evolución tecnológica y la adaptación del derecho procesal⁶⁴³.

Desde una perspectiva comparada, el modelo híbrido propuesto aspira a superar las limitaciones de ambas tradiciones procesales mediante la síntesis de sus fortalezas respectivas. Del *common law* adopta el énfasis en reglas claras de autenticación y estándares verificables de admisibilidad, que proporcionan seguridad jurídica y facilitan la predicción de resultados.

Del derecho continental incorpora el principio de valoración racional y la exigencia de motivación judicial, que garantizan la flexibilidad necesaria para adaptar la evaluación probatoria a las circunstancias específicas de cada caso⁶⁴⁴.

La viabilidad práctica del modelo propuesto depende de varios factores institucionales. En primer lugar, requiere el desarrollo de capacidades técnicas en el sistema judicial, incluyendo la formación de jueces en alfabetización digital avanzada, la creación de cuerpos de peritos especializados en análisis forense digital, y la disponibilidad de recursos tecnológicos adecuados para la realización de análisis complejos. En segundo lugar, demanda modificaciones legislativas específicas en los códigos procesales o reglas de evidencia, estableciendo los criterios de autenticación, las presunciones aplicables y los estándares de motivación. En tercer lugar, necesita el desarrollo jurisprudencial gradual que concrete y refine los principios generales mediante su aplicación a casos concretos⁶⁴⁵ (McPeak, 2021; Pfefferkorn, 2020).

⁶⁴² TYLER, T. R. (2006). *Why people obey the law*. Princeton University Press; FRIEDMAN, L. M. (2016). *Impact: How law affects behavior*. Harvard University Press.

⁶⁴³ HILDEBRANDT, M. (2015). *Smart technologies and the end(s) of law: Novel entanglements of law and technology*. Edward Elgar Publishing; WISCHMEYER, T., & RADEMACHER, T. (Eds.). (2020). *Regulating artificial intelligence*. Springer. <https://doi.org/10.1007/978-3-030-32361-5>.

⁶⁴⁴ TARUFFO, M. (2005). Conocimiento científico y estándares de prueba judicial. *Boletín Mexicano de Derecho Comparado*, 38(114), 1285-1312. <https://doi.org/10.22201/ijj.24484873e.2005.114.3770>; TWINING, W. (2006). *Rethinking evidence: Exploratory essays*. Cambridge University Press.

⁶⁴⁵ MCPEAK, A. (2021). Confronting the deepfake threat to trust in media, democratic deliberation, and evidence in court. *Vanderbilt Law Review*, 74(2), 445-504.

La implementación gradual del modelo puede seguir una estrategia de experimentación controlada, comenzando con su aplicación en tipos específicos de procedimientos donde la evidencia audiovisual resulta particularmente crítica y donde los recursos para análisis técnicos sofisticados justifican los costos involucrados. Los procedimientos penales que involucran delitos graves, los litigios civiles de alta cuantía y los procedimientos administrativos sancionadores constituyen contextos apropiados para esta experimentación inicial. La evaluación sistemática de la experiencia en estos contextos permitiría identificar ajustes necesarios antes de una implementación más amplia⁶⁴⁶

CONCLUSIONES

La irrupción de la inteligencia artificial generativa ha puesto al descubierto una vulnerabilidad estructural en los mecanismos de autenticación judicial. La posibilidad de fabricar registros audiovisuales indistinguibles de los reales no solo desestabiliza la confianza en la evidencia sensorial, sino que reconfigura la relación entre verdad, prueba y legitimidad institucional. La cuestión central no reside en la mera detección técnica de falsificaciones, sino en la revisión de los supuestos

epistemológicos que sostienen la valoración de la prueba audiovisual. El análisis evidencia que los sistemas jurídicos contemporáneos, anclados en paradigmas probatorios forjados en contextos tecnológicos distintos, carecen de herramientas conceptuales adecuadas para enfrentar una realidad donde la representación visual puede ser simulada con precisión casi perfecta.

El análisis realizado aporta una reformulación conceptual que trasciende las aproximaciones normativas convencionales. Propone un modelo híbrido de control judicial integral que combina verificación técnico-jurídica y valoración racional ampliada, lo cual redefine el equilibrio entre rigor epistémico y eficiencia procesal. Tal contribución se sustenta en la convergencia de dos tradiciones jurídicas y en la introducción de criterios verificables adaptados a la complejidad tecnológica. El resultado es una arquitectura procesal que recupera la legitimidad epistémica de la decisión judicial mediante la integración de fiabilidad técnica y razonabilidad argumentativa, desplazando el foco desde la prueba aislada hacia la coherencia sistémica del conjunto probatorio.

La traducción de las innovaciones teóricas en prácticas procesales implica transformaciones institucionales profundas. La

<https://scholarship.law.vanderbilt.edu/vlr/vol74/iss2/4/>; PFEFFERKORN, R. (2020). Deepfakes in the courtroom. *Boston University Journal of Science & Technology Law*, 29(1), 245-276. <https://papers.ssrn.com/sol3/papers.cfm?abs>

tract_id=3617472; TARUFFO, M. (2002). La prueba de los hechos. Trotta.

⁶⁴⁶ FERRAJOLI, L. (2011). *Principia iuris: Teoría del derecho y de la democracia*. Trotta; DWORKIN, R. (2011). *Justice for hedgehogs*. Harvard University Press.

alfabetización digital judicial emerge como un componente indispensable para el ejercicio efectivo del control epistémico. La creación de cuerpos periciales especializados y la asignación de recursos para tecnologías de detección se configuran como requisitos materiales de viabilidad. Además, la gobernanza adaptativa mediante comisiones interdisciplinarias de revisión normativa permite mantener actualizados los criterios procesales frente a la evolución tecnológica. Estas medidas consolidan un marco operativo capaz de garantizar la autenticidad y la integridad probatoria sin sacrificar la celeridad procesal ni la equidad de las decisiones.

La proyección del debate hacia futuras investigaciones invita a explorar la dimensión empírica del fenómeno. Se requiere una evaluación sistemática de la efectividad de los mecanismos de autenticación en distintos tipos de procesos y contextos jurisdiccionales. La colaboración entre juristas, informáticos, psicólogos y filósofos del derecho puede generar un corpus interdisciplinario más robusto que vincule la teoría de la prueba con la epistemología cognitiva y la ética de la inteligencia artificial. También se identifican líneas de investigación orientadas a la adaptación del modelo a los procedimientos penales, civiles y administrativos, donde los estándares de prueba y las exigencias de proporcionalidad varían significativamente.

El futuro del derecho procesal dependerá de su capacidad para reinventar sus fundamentos en un

entorno digital donde la percepción visual ya no garantiza veracidad. La transformación que se vislumbra no puede limitarse a un ajuste técnico, sino que debe asumir el carácter de una renovación filosófica del vínculo entre verdad, justicia y legitimidad. Reconocer que la autenticidad no es una propiedad inherente de la evidencia, sino una construcción institucional sostenida por procedimientos racionales implica asumir la tarea de repensar la función judicial desde una epistemología coherente, crítica y abierta al cambio. En esa dirección, la reconstrucción del proceso judicial frente a la amenaza de los *deepfakes* se configura como un compromiso ineludible con la preservación de la justicia en tiempos de simulación algorítmica.

DECLARACIÓN DE USO DE IA

Se usó el modelo de lenguaje GPT-5 de OpenAI para identificar y corregir errores de redacción y ortografía mediante el comando "detecta y corrige todos los errores de redacción y ortografía". Posteriormente, se revisó el texto obtenido para asegurar la fidelidad al estilo, contenido e intención del borrador original.

REFERENCIAS

- AJDER, H., PATRINI, G., CAVALLI, F., & CULLEN, L. (2019). The state of deepfakes: Landscape, threats, and impact. Deeptrace. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

- BRADFORD, A. (2020). The Brussels effect: How the European Union rules the world. Oxford University Press.
- CARNELUTTI, F. (1950). La prueba civil. Ediciones Depalma.
- CHESNEY, R., & CITRON, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753-1820. <https://doi.org/10.15779/Z38RV0D15J>
- CITRON, D. K., & CHESNEY, R. (2018). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98(1), 147-155.
- COUTURE, E. J. (1958). Fundamentos del derecho procesal civil. Ediciones Depalma.
- DELFINO, R. (2020). Deepfake porn and the impact to women and girls. University of San Francisco Law Research Paper. https://repository.uclawsf.edu/faculty_scholarship/761/
- DOBBE, T., METOUI, N., TRILLING, D., HELBERGER, N., & de VREESE, C. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *International Journal of Press/Politics*, 26(1), 69-91. <https://doi.org/10.1177/1940161220944364>
- DWORKIN, R. (2011). Justice for hedgehogs. Harvard University Press.
- FARID, H. (2019). Creating, using, misusing, and detecting deep fakes. *Journal of Online Trust and Safety*, 1(4), 1-16. <https://doi.org/10.54501/jots.v1i4.56>
- FERRAJOLI, L. (2011). Principia iuris: Teoría del derecho y de la democracia. Trotta.
- FERRER BELTRÁN, J. (2007). La valoración racional de la prueba. Marcial Pons.
- FLORIDI, L. (2018). Artificial intelligence, deepfakes and a future of ectypes. *Philosophy & Technology*, 31(3), 317-321. <https://doi.org/10.1007/s13347-018-0325-3>
- FLORIDI, L. (2021). The European legislation on AI: A brief analysis of its philosophical approach. *Philosophy & Technology*, 34(4), 215-222. <https://doi.org/10.1007/s13347-021-00460-9>
- FRIEDMAN, L. M. (2016). Impact: How law affects behavior. Harvard University Press.
- GASCÓN ABELLÁN, M. (2004). Los hechos en el derecho: Bases argumentales de la prueba. Marcial Pons.
- GRIMM, P., Capra, D., & Joseph, G. (2018). Evidence: A contemporary approach. West Academic Publishing.
- GÜERA, D., & DELP, E. J. (2018). Deepfake video detection using recurrent neural networks. *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, 1-6. <https://doi.org/10.1109/AVSS.2018.8639163>

- HAACK, S. (2014). Evidence matters: Science, proof, and truth in the law. Cambridge University Press.
- HARWELL, D. (2019). Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image'. The Washington Post. <https://www.washingtonpost.com/technology/2019/01/02/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/>
- HILDEBRANDT, M. (2015). Smart technologies and the end(s) of law: Novel entanglements of law and technology. Edward Elgar Publishing.
- HILDEBRANDT, M. (2020). Law for computer scientists and other folk. Oxford University Press.
- JASANOFF, S. (1995). Science at the bar: Law, science, and technology in America. Harvard University Press.
- KIETZMANN, J., LEE, L. W., MCCARTHY, I. P., & KIETZMANN, T. C. (2020). Deepfakes: Trick or treat? Business Horizons, 63(2), 135-146. <https://doi.org/10.1016/j.bushor.2019.11.006>
- MCPEAK, A. (2021). Confronting the deepfake threat to trust in media, democratic deliberation, and evidence in court. Vanderbilt Law Review, 74(2), 445-504. <https://scholarship.law.vanderbilt.edu/vlr/vol74/iss2/4/>
- NGUYEN, T. T., NGUYEN, C. M., NGUYEN, D. T., NGUYEN, D. T., & NAHAVANDI, S. (2019). Deep learning for deepfakes creation and detection: A survey. arXiv preprint arXiv:1909.11573. <https://doi.org/10.48550/arXiv.1909.11573>
- NIGHTINGALE, S. J., WADE, K. A., & WATSON, D. G. (2017). Can people identify original and manipulated photos of real-world scenes? Cognitive Research: Principles and Implications, 2(1), 1-21. <https://doi.org/10.1186/s41235-017-0067-2>
- PFEFFERKORN, R. (2020). Deepfakes in the courtroom. Boston University Journal of Science & Technology Law, 29(1), 245-276. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3617472
- TARUFFO, M. (2002). La prueba de los hechos. Trotta.
- TARUFFO, M. (2005). Conocimiento científico y estándares de prueba judicial. Boletín Mexicano de Derecho Comparado, 38(114), 1285-1312. <https://doi.org/10.22201/ij.24484873e.2005.114.3770>
- TARUFFO, M. (2009). La semplice verità: Il giudice e la costruzione dei fatti. Laterza.
- TERNOVSKI, J., Kalla, J., & Aronow, P. (2022). The negative consequences of informing voters about deepfakes: Evidence from two survey experiments. Journal of Online Trust and Safety, 1(2), 1-21. <https://doi.org/10.54501/jots.v1i2.36>

- TOLOSANA, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>
- TWINING, W. (2006). *Rethinking evidence: Exploratory essays*. Cambridge University Press.
- TYLER, T. R. (2006). *Why people obey the law*. Princeton University Press.
- TYLER, T. R., & HUO, Y. J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. Russell Sage Foundation.
- VACCARI, C., & CHADWICK, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1-13. <https://doi.org/10.1177/2056305120903408>
- VEALE, M., & BORGESIU, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97-112. <https://doi.org/10.9785/cr-2021-220402>
- VERDOLIVA, L. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5), 910-932. <https://doi.org/10.1109/JSTSP.2020.3002101>
- WESTERLUND, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11), 39-52. <https://doi.org/10.22215/timreview/1282>
- WISCHMEYER, T., & RADEMACHER, T. (Eds.). (2020). *Regulating artificial intelligence*. Springer. <https://doi.org/10.1007/978-3-030-32361-5>