Universidad de Guadalajara
DERECHO GLOBAL. ESTUDIOS SOBRE DERECHO Y JUSTICIA
Año 2024, Vol. IX. Número 27, Julio-Octubre, ISSN: 2448-5128 e-ISSN: 2448-5136
https://DOI.org/10.32870/dgedj.v9i27.754

FERNANDO RAMOS-ZAGA
Universidad Privada del Norte, Perú
fernandozaga@gmail.com

Deepfake: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial

Deepfake Technology: A Comprehensive Examination of Its Technological and Legal Implications in the Era of Artificial Intelligence

Cómo citar el artículo:

Ramos, F. (2024). Deepfake: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial. Derecho Global. Estudios sobre Derecho y Justicia, IX (27) https://10.32870/dgedj.v9i27.754 pp. 359-387

Recibido: 17/03/24 Aceptado: 08/05/24

Resumen

El surgimiento de la tecnología *deepfake* ha provocado un cambio trascendental en el mundo digital, exigiendo acciones proactivas tanto desde al ámbito académico como de los responsables políticos, debido a los múltiples desafíos que presenta a diferentes industrias. En ese sentido, el objetivo del presente artículo es analizar las implicancias tecnológicos y jurídicos en torno al *deepfake*, subrayando la necesidad imperiosa de que los marcos jurídicos se adapten a los retos que plantea el uso de la inteligencia artificial (IA). Para tal propósito, se llevó a cabo una revisión de la literatura especializada, concentrándose específicamente en los aspectos tecnológicos y los desafíos regulatorios que plantea la tecnología *deepfake*. Los hallazgos muestran la urgente necesidad de marcos jurídicos efectivos y acordes a los retos que plantea el vertiginoso avance tecnológico para hacer frente a los retos planteados a través del *deepfake* mediante la colaboración internacional para abordar estas cuestiones de manera integral. Se concluye que es de vital importancia contar con regulación que siga el ritmo de los avances tecnológicos para mitigar las posibles adversidades asociadas al uso de la tecnología *deepfake*.

Palabras clave

Ultrafalso, inteligencia artificial, legislación, tecnología, privacidad, seguridad, delitos, colaboración internacional.

Abstract

The emergence of deepfake technology has triggered a significant transformation in the digital landscape, necessitating proactive responses from academia and policymakers to address the diverse challenges it introduces across various sectors. This article aims to delve into the technological and legal dimensions of deepfake, stressing the critical importance of adapting legal frameworks to the evolving landscape of artificial intelligence (AI) applications. Through an extensive examination of scholarly sources, the analysis focuses on the technological intricacies and regulatory hurdles presented by deepfake technology. The research underscores the urgent requirement for robust legal structures that can effectively respond to the rapid advancements in technology, advocating for a collaborative approach on a global scale to tackle deepfake issues comprehensively. The study concludes that maintaining up-to-date regulations that parallel technological progress is essential in mitigating the potential negative impacts associated with the proliferation of deepfake technology.

Keywords

Deepfake, artificial intelligence, legislation, technology, privacy, security, crimes, international collaboration.

Sumario: I. Introducción. II. Inteligencia artificial, Big Data y aprendizaje automático. III. Deepfake. IV. Dilemas legales asociados a la tecnología Deepfake. 1. Manipulación de pruebas y extorsión. 2. Recopilación, intercambio y difusión de datos personales. 3. Generación de contenido explícito. 4.Violación de intimidad. 5. Difamación. 6. Infracción a la propiedad intelectual y artística. 7. Fraude mediante sistemas de información. 8. Propagación de información errónea durante procesos electorales. V. Estrategias a nivel internacional para contrarrestar los peligros que presenta la tecnología deepfake. VI. La urgencia de un enfoque jurídico integral: Desafíos en la regulación de la tecnología deepfake. VII. Conclusiones. Bibliografía.

I. Introducción

El advenimiento de la tecnología *deepfake* ha traído consigo un nuevo paradigma con importantes ramificaciones sociales en una era caracterizada por los rápidos avances de la inteligencia artificial (IA) y sus diversas aplicaciones (Karnouskos, 2020). La proliferación de los *deepfake*s, que implican la manipulación de contenidos audiovisuales mediante inteligencia artificial (IA) para lograr un efecto altamente realista, supone una amenaza significativa para múltiples industrias, tales como el cumplimiento de la ley, la política y los medios de comunicación (Pawelec, 2022). En ese contexto, la tecnología *deepfake* se está extendiendo por las plataformas digitales, lo que presenta una compleja interacción de consideraciones éticas, tecnológicas y legales (Li & Wan, 2023). Como resultado, los académicos y los responsables políticos requieren contar con una comprensión profunda de la cuestión y tomar medidas proactivas para abordarla.

Big Data proporciona vastos volúmenes de datos necesarios para entrenar modelos de IA, permitiendo que los algoritmos de ML y DL aprendan de grandes cantidades de ejemplos y variaciones (Nikolakopoulos et al., 2023). La IA, en su núcleo, habilita

la simulación de procesos cognitivos humanos, facilitando la creación de contenidos que imitan comportamientos y apariencias humanas de manera convincente (Staab, 2023). El aprendizaje automático, particularmente el aprendizaje profundo, emplea redes neuronales profundas que pueden identificar y replicar patrones complejos en datos visuales y auditivos.

Estas redes son capaces de generar imágenes y videos hiperrealistas mediante técnicas como las Generative Adversarial Networks (GANs), que son fundamentales en la producción de *deepfakes* (Naitali et al., 2023). Esta fusión de la creación de contenidos multimedia impulsada por la IA ha simplificado notablemente la producción de imágenes realistas pero generadas, difuminando los límites entre realidad y ficción como nunca antes. (Singh & Dhiman, 2023). El desarrollo de la tecnología *deepfake* subraya su naturaleza interdisciplinaria y su importancia académica, ya que cruza las ciencias de la computación, la psicología cognitiva y los estudios de los medios de comunicación, a la vez que se basa en los principios del aprendizaje automático y las redes neuronales (Whittaker et al., 2023).

El debate en torno a los *deepfake*s en el ámbito académico abarca múltiples campos y aborda diversas preocupaciones, que van desde los efectos psicológicos de los medios de comunicación engañosos hasta las consecuencias sociopolíticas de la difusión de información errónea (Hameleers et al., 2024). Los investigadores estudian activamente los intrincados procesos que intervienen en la creación y utilización de *deepfake*s, abordando cuestiones fundamentales relativas a la autenticidad, la legitimidad y la confianza en la era digital (Etienne, 2021). (Etienne, 2021). Por otra parte, la naturaleza en constante evolución de la tecnología *deepfake* pone de relieve la necesidad permanente de una investigación interdisciplinaria exhaustiva para dilucidar sus impactos multifacéticos en la sociedad en su conjunto (Vasist & Krishnan, 2022).

El uso generalizado de la tecnología *deepfake* supone un peligro importante para las instituciones democráticas mundiales y la credibilidad de los debates públicos (Berenzen, 2023), debido a que los *deepfake*s permiten crear contenido audiovisual muy realista, aunque pero completamente inexistente o parcialmente modificado, lo que provoca divisiones dentro de las comunidades, socava la confianza en las fuentes mediáticas establecidas e influye en la opinión pública a escala mundial

(Pawelec, 2022). Las consecuencias de la difusión incontrolada de *deepfake*s son amplias y significativas, y exigen la aplicación de normativas proactivas y un enfoque académico inmediato. Estas consecuencias abarcan no solo la difusión de información errónea por parte de individuos malintencionados, sino también el potencial de actividades ciberdelictivas y la explotación del porno de venganza.

La acuciante necesidad de comprender mejor la compleja relación entre los avances en inteligencia artificial (IA) y los marcos jurídicos en la lucha contra el fraude digital ha impulsado la realización del presente artículo que aborda los aspectos tecnológicos y jurídicos en torno al *deepfake*. Esta revisión bibliográfica tiene como objetivo consolidar los conocimientos existentes, identificar las áreas donde se necesita más investigación y ofrecer ideas sobre posibles vías para futuras investigaciones e intervenciones políticas mediante el análisis de trabajos académicos de diversas disciplinas. Asimismo, el presente estudio pretende concientizar a los responsables políticos, los profesionales del derecho y los desarrolladores de tecnología sobre la importancia de la adaptación proactiva frente a las amenazas en evolución impulsadas por la IA. Para ello, examina las capacidades tecnológicas de la tecnología *deepfake* y los retos normativos asociados a ella.

Mediante este trabajo, se resalta la necesidad de contar sistemas jurídicos sean proactivos y adaptables al panorama en rápida evolución de la inteligencia artificial y su aplicación en la generación de contenidos multimedia realistas, aunque completamente inexistentes o parcialmente modificados. De ese modo, se pretende fomentar una comprensión integral de las conexiones entre los avances tecnológicos, los marcos jurídicos y las implicancias sociales en el ámbito de la tecnología *deepfake*.

La presente investigación se realizó a través de una revisión de literatura pertinente en relación con las dimensiones tecnológicas y jurídicas de la tecnología *deepfake*. Se llevó a cabo una investigación bibliográfica exhaustiva utilizando bases de datos acreditadas como Google Scholar, Scopus y Web of Science, incorporando términos de búsqueda como "*deepfake* ", "artificial intelligence" y "legal frameworks". La búsqueda se restringió por fecha de publicación para abarcar trabajos de los últimos cinco años.

Los criterios de inclusión para este estudio abarcaron trabajos académicos, análisis jurídicos y documentos oficiales que analicen la tecnología *deepfake*, su impacto social y los enfoques normativos. Los criterios de exclusión consistieron en artículos de opinión y fuentes no revisadas por pares. Los artículos elegidos se evaluaron en función de su pertinencia, rigor e importancia, prestando especial atención a las perspectivas interdisciplinarias. La pertinencia se determinó considerando aquellos artículos que abordan directamente la intersección entre la inteligencia artificial, la generación de contenidos multimedia y los sistemas jurídicos. El rigor se evaluó en dos dimensiones principales. En primer lugar, el rigor con la teoría de la inteligencia artificial se clasificó como alto, medio o bajo, dependiendo de la profundidad y precisión con la que los artículos aplican y discuten las teorías y técnicas de inteligencia artificial. En segundo lugar, el rigor con la teoría del derecho se evaluó de manera similar, basándose en cómo los artículos integran y aplican conceptos jurídicos relevantes.

La revisión se centró en dar prioridad a los artículos que demostraban metodologías sólidas, y marcos teóricos cohesionados. En el proceso de síntesis de la literatura, se utilizaron técnicas de análisis temático y síntesis narrativa para mejorar la comprensión de los temas clave y ofrecer recomendaciones para futuras investigaciones e intervenciones mediante políticas públicas. Los resultados se integraron para presentar un examen exhaustivo de las dimensiones jurídicas y tecnológicas del *deepfake*, destacando la necesidad de una colaboración interdisciplinar y un ajuste proactivo para abordar eficazmente este problema emergente.

II. Inteligencia artificial, Big Data y aprendizaje automático

La definición de inteligencia artificial (IA) plantea que las máquinas sean capaces de imitar los procesos cognitivos humanos para mostrar comportamientos conscientes (Ng & Leung, 2020). Lograr esta simulación requiere una colaboración entre las capacidades de las máquinas y la inteligencia humana. El conocimiento empírico humano, derivado de percepciones y experiencias, influye en la formación de opiniones y procesos de pensamiento (Glikson & Woolley, 2020).

La intersección entre Big Data, inteligencia artificial (IA), aprendizaje automático (ML) y aprendizaje profundo (DL) es crucial para el desarrollo de *deepfakes*, ya que cada uno de estos componentes aporta capacidades esenciales para su creación y perfeccionamiento. La sinergia entre estas tecnologías no solo mejora la precisión y realismo de los *deepfakes*, sino que también acelera su desarrollo, planteando desafíos significativos en términos de ética, seguridad y regulación en el ámbito digital.

La importancia de los grandes conjuntos de datos reside en su capacidad para ser examinados mediante algoritmos de inteligencia artificial (IA), lo que facilita el reconocimiento de tendencias y la pronta aplicación de soluciones utilizando datos contextuales (Jagatheesaperumal et al., 2022). La integración de la IA con los macrodatos es crucial para manejar y comprender amplios conjuntos de datos; su ausencia haría inviable la creación de auténticos materiales *deepfake* (Gambín et al., 2024). Un componente integral de la IA es el aprendizaje automático o *deep learning*, mediante el cual una máquina adquiere conocimientos sustanciales sobre un tema concreto, los comprende y, posteriormente, aprovecha estos conocimientos para realizar funciones designadas (Khuat et al., 2022).

Por ende, la creación eficaz de contenidos *deepfake* auténticos depende en gran medida de la interdependencia de amplios conjuntos de datos y técnicas avanzadas de aprendizaje automático. La sinergia entre estos componentes eleva la precisión y la autenticidad, enfatizando la calidad experiencial de los datos (Waseem et al., 2023). En esencia, la fusión de la inteligencia artificial, los vastos recursos de datos y el sofisticado aprendizaje automático significa una transición revolucionaria en la que los sistemas informáticos se esfuerzan por emular y mejorar las funciones cognitivas similares al comportamiento humano. (Tariq et al., 2023).

III. Deepfake

La aparición del término "*deepfake*", como combinación de "*deep*" (profundo) y "*fake*" (falso), se define como el uso de técnicas de aprendizaje profundo para fabricar contenidos mediáticos engañosos. Este término engloba tanto el proceso de creación como los medios manipulados resultantes (Heidari et al., 2023). La

tecnología *deepfake* permite alterar el contenido de los medios de comunicación sin interacción directa con el material original, lo que implica esencialmente la sustitución del rostro de una persona por el de otra en material mediático auténtico (Mukta et al., 2023).

Es esencial aclarar que las manipulaciones van más allá de la sustitución facial, pues abarcan todo el físico, incluidas partes específicas del cuerpo asociadas a un individuo concreto (Ayers, 2021). Asimismo, la tecnología *deepfake* trasciende los medios visuales para permitir manipulaciones de audio, especialmente en contextos de vídeo. Esta tecnología puede alterar auténticamente un diálogo hablado y generar contenidos en los que aparezcan personas pronunciando declaraciones que en realidad nunca hicieron (Farid, 2022).

IV. Dilemas legales asociados a la tecnología *Deepfake*

La utilización de la tecnología *deepfake* presenta importantes dilemas éticos y jurídicos, a pesar de sus posibles ventajas (Diakopoulos & Johnson, 2021). Principios jurídicos fundamentales como la protección del bienestar físico y espiritual, el derecho a la intimidad y el respeto de la vida personal y familiar están en peligro (Rizzica, 2021). En la misma línea, elementos clave como la libertad de búsqueda artística y científica, junto con el derecho a expresar ideas y puntos de vista individuales, cobran una importancia crucial en el contexto del *deepfake* (Mammadzada, 2021). En particular, la maleabilidad de las percepciones humanas plantea una grave amenaza. Por ejemplo, entidades malintencionadas podrían crear vídeos engañosos utilizando la imagen de alguna persona en contextos inapropiados, lo que podría dar lugar a extorsión (Greenough, 2022).

Las implicaciones de la tecnología *deepfake* van más allá del ámbito personal, adentrándose en la esfera de la seguridad nacional, como ejemplifica un caso conmovedor de un vídeo de Jordan Peele en el que aparece la imagen falsa de Barack Obama (Bode et al., 2021).Estos vídeos *deepfake* tienen el potencial de influir en la dinámica social y las relaciones internacionales, alertando a los organismos gubernamentales sobre cuestiones pertinentes. Reconociendo la gravedad de la situación, el Comité Permanente Selecto de Inteligencia de la

Cámara de Representantes de Estados Unidos destaca la amenaza que supone el uso indebido de los sistemas de información en la creación de contenidos *deepfake*, que puede poner en peligro tanto a las personas como a la seguridad nacional (Pantserev, 2020).

Las repercusiones de la tecnología *deepfake* en el ámbito jurídico abarcan una serie de implicancias complejas, pues si bien pueden utilizarse sistemas de información en su desarrollo, el mero acto de producir contenido *deepfake* no necesariamente constituiría necesariamente un delito (Montasari, 2024). Por ese motivo, durante la determinación de la culpabilidad, la intencionalidad desempeña un papel fundamental (Kirchengast, 2020). Por otro lado, el acto de generar material *deepfake* requiere un análisis de intencionalidad para explotar este contenido con objetivos de cometer fraude (Busacca & Monaca, 2023).

1. Manipulación de pruebas y extorsión

En el ámbito de los delitos existe una conexión con la difamación y su compleja relación con la tecnología *deepfake*. En concreto, la generación y difusión de material *deepfake* puede servir como herramienta para manipular pruebas, lo cual permitiría incriminar erróneamente a una persona por un presunto delito (Rizzica, 2021). En ese sentido, la aparición de la tecnología *deepfake* presenta obstáculos en la presentación de pruebas digitales, ya que permite a personas con conocimientos tecnológicos crear falsificaciones casi indistinguibles, lo que socava la credibilidad de las pruebas. (Mammadzada, 2021). Por ende, este avance tecnológico suscita preocupaciones, ya que facilitaría la manipulación de pruebas de naturaleza audiovisual con la finalidad de inducir percepciones erróneas respecto a la comisión de conductas ilícitas.

La extorsión consiste en la amenaza de revelar o relacionar detalles desfavorables a la personalidad de una persona para obtener el control sobre ella o beneficios indebidos. Una importante evolución de este delito consiste en difundir información falsa que parece auténtica, lo que permite a los autores explotar la tecnología *deepfake* para fabricar contenidos verosímiles pero engañosos que dañan la reputación de una persona (Mustak et al., 2023). En consecuencia, la aparición de aplicaciones *deepfake* introduce complejidades y retos en el marco jurídico, lo que exige un examen meticuloso y estrategias preventivas.

2. Recopilación, intercambio y difusión de datos personales

La relación entre la disponibilidad de grandes volúmenes de datos y la credibilidad de los resultados permite generar contenido con mayor índice de autenticidad (Lu & Ebrahimi, 2024). Concretamente, un conjunto de datos faciales diverso y extenso es vital para manipular elementos faciales en contenidos visuales. Por tanto, es esencial comprender que los atributos faciales, visuales o vocales distintivos utilizados en el aprendizaje profundo constituyen información personal, definida como datos relativos a un individuo reconocido o identificable (Heidari et al., 2023). La obtención no autorizada de datos personales puede tener repercusiones penales asociadas a delitos informáticos de acuerdo a las legislaciones vigentes de cada país (Mekkawi, 2023).

Las implicaciones jurídicas van más allá de la obtención no autorizada de datos y abarcan acciones posteriores que implican la alteración de datos y la difusión de imágenes. El tratamiento de datos personales implica la gestión de información personal a través de diversas actividades, como la obtención, registro, almacenamiento, modificación, divulgación, transferencia y suministro de datos, ya se ejecuten de forma total o parcialmente automática o dentro de un sistema de datos estructurado (Mekkawi, 2023).

3. Generación de contenido explícito

El avance de la tecnología *deepfake* permite alterar imágenes y vídeos para crear contenidos engañosos en los que se representa a personas en contextos inapropiados, el cual está dirigido sobre todo a las mujeres, plantea dilemas éticos, especialmente en relación con las posibles implicaciones de las disposiciones relacionadas con la obscenidad y la pornografía (O'Halloran, 2021).

La utilización de la tecnología *deepfake* para mezclar rasgos físicos de personas auténticas con contenidos inapropiados plantea un problema importante en el material audiovisual para adultos. La principal preocupación gira en torno a la posible ilegalidad del contenido resultante, especialmente cuando se incorpora la imagen de una persona que no participa originalmente en el vídeo (Popova, 2020). Por ejemplo, sustituir el rostro de la persona A por el de la persona B en una escena explícita induce a error a los espectadores, haciéndoles creer que la persona A participó en dicha actividad.

Asimismo, la legislación regional prohíbe la participación de menores de edad en la producción de contenido pornográfico (Pascale, 2023). Por tanto, es fundamental distinguir entre la producción de contenidos *deepfake* de temática adulta y su accesibilidad para los menores. Si bien la creación de este tipo de material puede no considerarse intrínsecamente obscena, su ilegalidad surge en su difusión, venta o exhibición de forma accesible a menores de edad (Rizzica, 2021). Por lo tanto, el examen jurídico debe abarcar no sólo el proceso de creación, sino también un espectro más amplio, el cual comprende la distribución y el acceso del contenido.

4.Violación de intimidad

La violación de la intimidad de una persona mediante la distribución no autorizada de imágenes o grabaciones de audio privadas es un delito punible en el ámbito regional. No obstante, la creciente utilización de la tecnología *deepfake* facilita este tipo de violaciones al permitir la creación y difusión de contenidos mediáticos fabricados (Pantserev, 2020). Este avance supone una amenaza significativa, ya que permite la generación y circulación de material explícito falsificado en el que participan personas que no tienen ninguna relación real con él.

El acto de producir y compartir material sexualmente explícito utilizando métodos *deepfake* supone una violación de la intimidad, lo que lo convierte en un delito penal. Esta violación se produce cuando se manipulan imágenes o vídeos de personas a las que no se suele asociar con contenidos explícitos (Okolie, 2023). Es importante señalar que, aunque en el material difundido aparezcan los cuerpos de las personas sin que sus rostros sean claramente visibles, se sigue produciendo una violación de la intimidad.

5. Difamación

En el contexto jurídico, la difamación se define como la acción que menoscaba la reputación, la integridad y la posición social de una persona al asociarla con un determinado hecho o situación (Jinana, 2020). Esta transgresión va más allá de las formas convencionales de interacción y abarca mensajes escritos o audiovisuales, dirigidos específicamente a la víctima. El uso indebido de la tecnología *deepfake* para producir contenidos en los que aparezcan personas que no existen o son retratadas de forma inexacta en comparación con la realidad puede suponer una amenaza para la reputación, la integridad y la estima de las personas implicadas (Mammadzada, 2021).

Al respecto, se puede considerar a la persona A, quién es respetado en su círculo social. Cuando individuos malintencionados obtienen fotografías de A con un atuendo formal, combinan estas imágenes con la imagen de una persona no relacionada B, la cual es muy conocida por su extenso prontuariado. Esta manipulación visual retrata a la persona de manera deshonesta, perjudicando la percepción del buen nombre y prestigio de A. En este escenario, la creación y difusión de contenidos *deepfake* se asemeja mucho a la difamación, ya que implica un esfuerzo calculado para desprestigiar a un individuo en la sociedad.

6. Infracción a la propiedad intelectual y artística

Las infracciones relativas a la propiedad intelectual comprenden la vulneración contra los derechos éticos, financieros o conexos vinculados a obras creativas e intelectuales. Estas transgresiones se manifiestan como alteraciones no autorizadas de creaciones, exhibiciones, grabaciones o cualquier otro producto intelectual o creativo que incorpore las características distintivas de sus creadores, especialmente en campos como la ciencia, la literatura, la música, las artes visuales o el cine. (Pavis, 2021). Por ejemplo, el uso indebido de la tecnología *deepfake* para manipular la escena de una película constituye una infracción si se lleva a cabo sin el consentimiento explícito por escrito de los titulares de los derechos. Por consiguiente, las modificaciones no autorizadas de obras artísticas suponen infracciones sustanciales de los derechos de propiedad intelectual, lo que acentúa la compleja interacción entre las estructuras jurídicas que salvaguardan las actividades creativas e intelectuales y los avances tecnológicos (Schick, 2020).

7. Fraude mediante sistemas de información

El contenido *deepfake* comprende material alterado en formato de audio, imagen y vídeo. La producción de estos contenidos depende del uso de un Sistema de Gestión de Contenidos para fabricarlos y de la explotación de otro sistema de información para hacerlos circular, lo que amplifica el potencial de engaño. Este complejo procedimiento implica el empleo de tecnologías sofisticadas, como algoritmos de aprendizaje automático, para imitar o sustituir de forma realista componentes originales dentro de los medios de comunicación, lo que contribuye a la naturaleza compleja y engañosa de los *deepfakes*. El uso simultáneo de sistemas de información

pone de relieve la complejidad del proceso de producción de *deepfakes*, en el que se entrecruzan la manipulación y la distribución de contenidos, lo que crea dificultades para detectar y prevenir estas actividades engañosas (Vasist & Krishnan, 2022).

El uso de un Sistema de Gestión de Contenidos puede facilitar actividades fraudulentas mediante la creación de contenidos *deepfake*. Asimismo, cualquier medio que se utilice para difundir contenidos engañoso, puede servir como herramienta para actividades delictivas (Mustak et al., 2023). Por ejemplo, un individuo que genera material *deepfake* en un ordenador personal y lo comparte en una plataforma para su difusión, hace uso de este medio como herramienta para perpetrar un fraude.

8. Propagación de información errónea durante procesos electorales

El reto que supone el *deepfake* para comprometer la integridad electoral es importante, ya que sirve como potente herramienta para distorsionar y socavar los procedimientos democráticos, lo que supone una grave amenaza para la estabilidad política. El uso estratégico de material *deepfake* para influir en los resultados electorales plantea cuestiones alarmantes, que sobrepasan los límites del diálogo político ético. Estos esfuerzos implican la circulación de contenidos alterados que no sólo erosionan la fiabilidad de los procedimientos electorales, sino que también representan una forma de propaganda sofisticada. De ese modo, las personas implicadas en la generación de material *deepfake* pretenden sacar provecho de las debilidades del sistema democrático, vulnerando la confianza pública y fomentando la discordia mediante el uso de tecnologías de vanguardia con la finalidad de construir narrativas creíbles y difundir afirmaciones infundadas (Schick, 2020).

V. Estrategias a nivel internacional para contrarrestar los peligros que presenta la tecnología *deepfake*

La creciente preocupación por el impacto de la tecnología *deepfake* ha puesto de relieve la necesidad de un examen detallado en el ámbito del derecho penal y la política para evaluar los avances jurídicos como paso fundamental hacia la aplicación de futuras medidas legales destinadas a mitigar los riesgos inherentes

asociados a la tecnología *deepfake*. En ese sentido, el enfoque principal de este análisis gira en torno a las legislaciones que siguen evolucionando, tanto en Estados Unidos como la Unión Europea.

Avances regulatorios en Estados Unidos para abordar los desafíos de la tecnología deepfake

En la batalla legal global para combatir los riesgos asociados a la tecnología *deepfake*, Estados Unidos se ha situado como uno de los países pioneros en este ámbito. En el Estado de Virginia, la Virginia Assembly Bill 602 prohíbe la venta o distribución de contenido *deepfake*, lo que abarca el uso no autorizado de imágenes de particulares. Inicialmente, la difusión de imágenes auténticas y lascivas con fines de coacción o acoso se consideraba ilegal según las disposiciones del Código de Procedimiento Penal de Virginia sobre delitos de obscenidad. Posteriormente, el 18 de marzo de 2019, se implementó un ajuste para abarcar contenido explícito manipulado, lo que refleja un reconocimiento temprano de las amenazas en evolución que plantea la tecnología *deepfake* (Kugler & Pace, 2021).

En el estado de Texas se implementó el 1 de septiembre de 2019, para hacer frente a la amenaza de los videos *deepfake* en las elecciones, la Texas Bill 751, una enmienda al Código Electoral de Texas, que prohíbe la creación y difusión de videos *deepfake* dentro de los 30 días de una elección, con el objetivo de dañar a los candidatos o manipular los resultados electorales. Los videos *deepfake* en este contexto se refieren a videos engañosos que retratan a individuos participando en acciones que en realidad no realizaron. Del mismo modo, el 22 de enero de 2019, Massachusetts presentó el proyecto de ley H.3366, que se centra en la regulación de la creación y distribución de contenido *deepfake*, especialmente dirigido a aquellos que producen o difunden *deepfake*s con fines maliciosos (Langa, 2021).

En California se presentó el proyecto de ley 730 de la Asamblea (Berman) para hacer frente a la creciente amenaza que suponen los *deepfake*s. El proyecto de ley establece la prohibición de difundir, exhibir o compartir contenidos *deepfake*. En particular, la legislación prohíbe el uso de material *deepfake* para engañar a los votantes o dañar la reputación de los candidatos en las contiendas electorales (Lussier, 2022). El 12 de junio de 2019, a nivel federal, se presentó la *DEEP FAKES*

Accountability Act, una iniciativa legislativa que propone la creación de un registro de identidades fabricadas mediante tecnología avanzada y tiene como objetivo combatir la propagación de la desinformación facilitada por la tecnología *deepfake* (Kugler & Pace, 2021).

La *DEEP FAKES Accountability Act* impide la creación de contenidos *deepfake* con la intención de difamar o desacreditar mediante representaciones de desnudos o actividades sexuales. Tipifica como delito las acciones que instiguen a la violencia, interfieran en las funciones oficiales, manipulen los debates sobre políticas públicas o perturben los procesos electorales. Asimismo, la legislación distingue entre grabaciones inalteradas y las modificadas sustancialmente mediante técnicas de *deepfake*. En algunos casos, puede haber excepciones para el uso de material *deepfake* ante circunstancias específicas, como medidas de seguridad nacional o pública (Langa, 2021).

Por otro lado, la propuesta legislativa conocida como *Deepfake* Task Force Act recomienda crear un equipo especializado que se ocupe de identificar y evaluar los riesgos vinculados a la tecnología *deepfake*, así como de formular estrategias y políticas para mitigarlos. La propuesta es un componente fundamental de iniciativas de seguridad más amplias destinadas a salvaguardar los intereses nacionales de los peligros que plantea la tecnología *deepfake* y a ofrecer recursos legales a las víctimas afectadas por *deepfake*s maliciosos. La propuesta se encuentra actualmente en proceso legislativo en los Estados Unidos, en el marco del 117º Congreso (2021-2022), y debe ser revisada por el Comité de Seguridad Nacional de la Cámara de Representantes. (Congreso de Estados Unidos, 2022).

Marco regulatorio en la Unión Europea ante los desafíos de la Inteligencia Artificial y los deepfakes

En febrero de 2019, el Parlamento Europeo aprobó la Directiva (UE) 2018/958, la cual se centra principalmente en abordar cuestiones relacionadas con *deepfake*s dentro del sistema de justicia penal (Duminica & Ilie, 2023). En colaboración con la Alianza Europea de IA, se plantean directrices que fomenten una estrategia europea cohesiva sobre el desarrollo de la IA. La motivación de esta directiva se debe a la urgente necesidad de un enfoque consolidado a escala de la UE para contrarrestar

las considerables inversiones en IA de los rivales mundiales, especialmente China y Estados Unidos (Leiser, 2022). Cabe destacar en esta Directiva su énfasis en abordar las amenazas potenciales para la democracia y los derechos fundamentales que plantea el uso malintencionado de la IA, en particular en lo relativo a la seguridad digital, el bienestar público y la autonomía individual.

Esta iniciativa construyó un marco que penaliza las prácticas que implican manipulación de la percepción con el fin de contrarrestar los efectos negativos de la IA, en particular en los casos en que el contenido personalizado distorsiona la realidad y tiene efectos desfavorables (Mahashreshty, 2023). Del mismo modo, las medidas legislativas en Europa se centran en la lucha contra las estrategias engañosas, concretamente en el contexto de la influencia en las elecciones mediante la creación de contenidos falsos.

Adicionalmente, el Parlamento Europeo y el Consejo presentaron el 21 de abril de 2021 la propuesta de Reglamento de Inteligencia Artificial, con el objetivo de crear normas uniformes en el marco de la Ley de Inteligencia Artificial. Esta iniciativa establece responsabilidades claras, justas y explícitas a los proveedores y usuarios de sistemas de IA para garantizar la seguridad y el cumplimiento de la normativa vigente que protege los derechos fundamentales. El documento establece directrices basadas en el principio de transparencia, especialmente en lo relativo a las tecnologías *deepfake*s y chatbot (Tahraoui et al., 2023).

Los retos que plantea la información errónea se destacan en la Guía de la Comisión Europea para mejorar la normativa sobre desinformación, especialmente en relación con la crisis de COVID-19. En el documento se abordan diversos modos de manipulación, incluidas amenazas emergentes como la tecnología *deepfake*, exigiendo estrategias eficaces para combatir estas prácticas engañosas. En medio de este amplio panorama normativo, la Unión Europea está formulando proactivamente medidas estrictas para protegerse contra los efectos adversos de la inteligencia artificial y la información falsa (Bayer et al., 2021).

De acuerdo a Casals (2023), la Ley de Inteligencia Artificial (IA) de la Unión Europea, también conocida como Ley IA, es la primera normativa exhaustiva sobre IA implementada por un regulador de gran relevancia a nivel mundial.

Propuesta inicialmente en abril de 2021, la ley clasifica las aplicaciones de IA en tres categorías de riesgo: aplicaciones prohibidas, como los sistemas de puntuación social administrados por el gobierno; aplicaciones de alto riesgo, como las herramientas de selección de CV, que están sujetas a requisitos legales específicos y deben ser evaluadas antes de su comercialización y durante su ciclo de vida; y aplicaciones de bajo riesgo, que quedan en gran medida sin regular. La ley busca garantizar que los sistemas de IA utilizados en la UE sean seguros, transparentes, trazables, no discriminatorios y respetuosos con el medio ambiente. Además, establece requisitos de transparencia para la IA generativa, como ChatGPT, que debe revelar que el contenido ha sido generado por IA y evitar la generación de contenidos ilegales. Tras largas negociaciones, el Parlamento Europeo y el Consejo de la UE alcanzaron un acuerdo provisional sobre la ley en 2023, estableciendo un marco regulador robusto que fomenta la innovación tecnológica guiada por principios éticos y legales en la UE.

VI. La urgencia de un enfoque jurídico integral: Desafíos en la regulación de la tecnología *deepfake*

La propagación global de la tecnología *deepfake* es evidente a través de diversos casos en todo el mundo. Tal es el caso ocurrido en México, donde un estudiante de educación superior identificado como Diego "N", ha sido acusado de poseer material pornográfico en su iPad, el cual contenía aproximadamente 160,000 imágenes íntimas, videos y fotos, tanto reales como manipuladas, de alrededor de mil mujeres, varias de las cuales eran estudiantes de su institución educativa. La investigación reveló que Diego "N" había utilizado varias aplicaciones para descargar y almacenar la información, lo que llevó a las autoridades a investigar posibles casos adicionales. Este caso ha generado preocupación entre padres de familia y autoridades educativas, ya que representa una violación a la privacidad y seguridad de las mujeres involucradas (El País México, 2023).

En Chorrillos, Lima, Perú, se produjo un incidente en el que unos estudiantes manipularon indebidamente las imágenes de sus compañeros utilizando inteligencia artificial (IA) y luego las comercializaron como material pornográfico. En consecuencia, las autoridades han iniciado investigaciones sobre la utilización

de la tecnología *deepfake* para la producción y difusión de contenidos explícitos. George's College, la institución implicada en el asunto, ha especificado que los estudiantes bajo sospecha fueron suspendidos temporalmente a la espera de una investigación exhaustiva. Este caso subraya la necesidad de contar con regulaciones y repercusiones bien definidas en relación con el uso indebido de la tecnología *deepfake*, especialmente en los casos que implican engaño y daño a personas vulnerables como los menores de edad. (Perú21, 2023).

En Almendralejo, Badajoz, Extremadura, España, numerosas jóvenes se han convertido en víctimas de imágenes de desnudos generadas por inteligencia artificial y distribuidas entre sus compañeros a través de teléfonos inteligentes. Estas imágenes procedían de fotos de las estudiantes obtenidas de sus perfiles de las redes sociales, que fueron manipuladas utilizando un software basado en IA para producir versiones explícitas. Este preocupante suceso ha dado lugar a investigaciones sobre la producción y difusión de estas imágenes falsas, que han revelado la participación de al menos once estudiantes en el intercambio de contenidos a través de aplicaciones de mensajería como WhatsApp y Telegram. Asimismo, hay acusaciones de intentos de extorsión contra ciertas víctimas, amenazándolas con exponer estas imágenes falsificadas a sus familias (BBC News, 2023). En consecuencia, el impacto de esta situación ha infundido temor y ansiedad entre la población femenina joven, haciéndola sentir expuesta y confinada en sus residencias.

Por ende, a falta de una legislación específica que regule las consecuencias jurídicas de las manipulaciones *deepfake*, es evidente que abordar esta cuestión no es actualmente una preocupación primordial en el ordenamiento jurídico regional. Sin embargo, a partir de modelos internacionales que han implementado estructuras legales, tal como es el caso de Estados Unidos y Unión Europea, se hace cada vez más evidente el imperativo de adoptar medidas legislativas en este ámbito. Por lo tanto, es esencial explorar posibles disposiciones penales en el contexto regional, específicamente en relación con los delitos en los que el contenido *deepfake* desempeñan un rol gravitante.

En la actualidad, el marco jurídico vigente aborda los delitos relacionados sin tener en consideración el contenido *deepfake, lo cual* requiere ampliar el ámbito de

aplicación de la legislación vigente. No obstante, puede resultar esencial establecer protocolos jurídicos específicos para los casos en que los contenidos *deepfake* se utilicen en actividades delictivas, en particular en casos como difamación, fabricación de pruebas, extorsión, fraude y vulneración de la propiedad intelectual y creativa, teniendo en cuenta que estas transgresiones podrían ser agravadas cuando están vinculadas a la generación, utilización, exposición o difusión de material *deepfake*.

A pesar de los avances en la detección de contenidos *deepfake*, lograr una fiabilidad total del contenido audiovisual continúa siendo un reto, lo cual perpetúa la naturaleza engañosa de los *deepfake*s, que exigen un escrutinio minucioso. Esta dificultad persistente se ve acentuada por los continuos avances en las tecnologías de *deepfake*, que disminuyen la eficacia de los métodos de detección tradicionales. En consecuencia, el riesgo de que los deepfakes induzcan a error a los espectadores, distorsionen las percepciones y comprometan la exactitud de los datos visuales y auditivos sigue siendo un problema crítico. Para hacer frente a este reto es necesario explorar y mejorar continuamente técnicas de detección sofisticadas, empleando una estrategia interdisciplinar que combine innovaciones tecnológicas, puntos de vista éticos y marcos jurídicos para proteger contra las repercusiones negativas del material audiovisual engañoso en la confianza del público, la fiabilidad de la información y los sistemas democráticos.

La existencia de lagunas jurídicas en relación a la recopilación, difusión y obtención no autorizada de datos personales mediante el uso de tecnología *deepfake* presenta cierta complejidad debido a la existencia de una entidad compuesta que se presenta como auténtica pero que es fundamentalmente falsa. Esta manipulación va más allá de la mera posesión o transmisión de datos legítimos. Por lo tanto, es aconsejable adoptar una definición exhaustiva de las actividades ilícitas vinculadas al uso indebido de la tecnología *deepfake*, que abarque acciones como la alteración de la imagen o la voz de otra persona, el empleo de la identidad de otro individuo en contenidos fabricados y la circulación o exposición encubiertas de ese material engañoso. Para tal fin, es esencial considerar la nota técnica del BID titulada "Adopción ética y responsable de la inteligencia artificial en América Latina y el Caribe" (Pombo et al., 2020). En el mismo sentido, la "Recomendación sobre la Ética de la Inteligencia Artificial" de la UNESCO de 2021 ofrece un marco

integral para guiar a los Estados miembros en el desarrollo y uso responsable de la IA (UNESCO, 2022). Además, los "Principios de la OCDE sobre Inteligencia Artificial" de 2019 proporcionan un marco robusto para el desarrollo y uso responsable de la IA (OCDE, 2019).

El examen de criminalidad respecto a la producción de contenido explícito basado en *deepfake,* plantea diversas aristas. En concreto, el acto de generar material *deepfake* sin el consentimiento de las personas representadas plantea dilemas éticos. No obstante, la perspectiva jurídica del delito de exhibiciones o publicaciones obscenas no aborda de forma proactiva la cuestión fundamental del uso de esta tecnología, pues los contenidos *deepfake* suscitan preocupaciones en relación con la privacidad, ya que amalgaman imágenes suplementarias y no originales con datos personales, sobre todo en contextos sexualmente sugerentes. Por ende, el concepto de definir el ámbito de ilicitud respecto a la alteración o uso indebido de la imagen o la voz de alguna persona sin permiso se alinea con la salvaguarda de la información personal.

La gravedad de las violaciones de la intimidad, especialmente en los casos de material sexualmente explícito, resalta la necesidad de incorporar un componente punitivo en el tratamiento de actividades relacionadas con la producción de contenido *deepfake*. Del mismo modo, cuando dicho contenido atenta contra el honor, la dignidad o la reputación de una persona, se pasa al ámbito de la difamación y se justifica una intervención desde la persecución del delito. Por tanto, es imperativo realizar un análisis meticuloso e interdisciplinario debido a su alto nivel de sofisticación.

VII. Conclusiones

El vertiginoso avance de la tecnología *deepfake* plantea importantes retos éticos y jurídicos que exigen la pronta actuación de la comunidad jurídica. Los *deepfake*s tienen la capacidad de generar contenido multimedia muy realista pero totalmente falso, lo que supone una intromisión en la intimidad personal y una amenaza para la seguridad nacional. Esto subraya la importancia de que los marcos legislativos sean proactivos a la hora de adaptarse a las tecnologías emergentes e incorporar nuevas disposiciones legales para abordar los riesgos inherentes asociados a la tecnología

deepfake. En consecuencia, las leyes deben demostrar flexibilidad en respuesta a la evolución del panorama tecnológico, al tiempo que aplican medidas para mitigar los posibles daños derivados del uso indebido de las *deepfake*s.

Los marcos jurídicos existentes se enfrentan a un reto importante a la hora de abordar eficazmente los complejos problemas que plantean las *deepfake*s. Este contenido alterado digitalmente es sofisticado y difícil de detectar, lo que crea barreras únicas dentro del marco jurídico establecido. Para combatir este problema, es crucial desarrollar nuevas definiciones y categorizaciones legales que estén específicamente diseñadas para tratar los delitos relacionados con los *deepfake*s. Estos marcos jurídicos deben tener en cuenta la compleja naturaleza de estos delitos, incluida la manipulación engañosa y fraudulenta de contenidos visuales y auditivos. Asimismo, es evidente que deben aplicarse penas más estrictas para los delitos relacionados con los *deepfake*s, con el fin de disuadir a los actores maliciosos y salvaguardar a las personas y a la sociedad de las posibles consecuencias del uso indebido de los *deepfake*s.

La colaboración internacional es crucial para afrontar con eficacia los retos jurídicos que plantean los *deepfake*s. Con el uso generalizado de la tecnología *deepfake* y la naturaleza global de los medios digitales, es imperativo fomentar la cooperación internacional con el fin de establecer marcos jurídicos sólidos capaces de hacer frente a diversos tipos de delitos relacionados con los *deepfake*s. Este enfoque de colaboración facilitaría el intercambio de ideas y estrategias, lo cual promovería la armonización de las normas jurídicas y apoyaría los esfuerzos unificados a nivel global para mitigar los efectos nocivos del *deepfake*.

El contenido *deepfake* plantea diversos retos que pueden abordarse eficazmente adoptando un enfoque multidisciplinar que integre estrategias de los campos de la tecnología, el derecho y la educación. Una parte integral de este enfoque es el uso de algoritmos avanzados de inteligencia artificial y aprendizaje automático para identificar y marcar rápidamente los contenidos manipulados, facilitando así la detección automática de *deepfake*s. Además de las soluciones tecnológicas, la sensibilización del público a través de iniciativas educativas desempeña un papel crucial a la hora de informar a las personas sobre la existencia y las posibles consecuencias del *deepfake*. De ese modo, se fomenta la capacidad de evaluar críticamente las fuentes y contenidos de los medios de comunicación.

La eventual exacerbación del fenómeno del *deepfake* y su eventual transformación en una "tragedia irreversible", constituye la razón subyacente que exige una urgente respuesta desde el ámbito jurídico y social que fomente la innovación responsable en inteligencia artificial (IA) y tecnologías de medios digitales, al tiempo que penalice a quienes se dedican a la creación y distribución maliciosa de *deepfake*s. La respuesta jurídica a la tecnología *deepfake* debe adaptarse a su rápida evolución y, simultáneamente, proteger a la sociedad de su uso indebido. Este enfoque garantizará la autenticidad de los contenidos digitales y protegerá a las personas y organizaciones de los efectos negativos del contenido manipulado.

Las investigaciones futuras deberían dar prioridad al desarrollo de un marco jurídico integral que permita abordar los retos únicos que plantea la tecnología *deepfake* con la finalidad de identificar las lagunas y los puntos débiles de los marcos jurídicos existentes en las distintas jurisdicciones a la hora de abordar los delitos relacionados con la *deepfake*. Para tal fin, se puede realizar un análisis exhaustivo de derecho comparado con el objetivo de proponer una norma jurídica unificada que pueda ser aplicada en el ámbito regional. Asimismo, esta investigación podría explorar el delicado equilibrio entre la prevención del uso indebido de *deepfake*s y la salvaguarda de la libertad y la innovación, para lo cual debe contar con propuestas legislativas que respetuosas con los derechos fundamentales. Los resultados de este estudio desempeñarán un papel crucial en el avance de respuestas jurídicas más sólidas y eficaces al campo de la tecnología *deepfake*, que evoluciona rápidamente.

Es esencial contar con profesionales del Derecho que reciban formación adecuada para emplear herramientas tecnológicas diseñadas para detectar *deepfake*s. Esta formación contribuirá significativamente a mejorar la identificación de contenido manipulado. Asimismo, los marcos jurídicos deben actualizarse para incluir disposiciones específicas sobre los delitos relacionados con los *deepfake*s. De este modo, se garantiza que las leyes puedan abordar eficazmente las complejidades de este tipo de situaciones. La colaboración entre abogados, instituciones educativas, actores tecnológicos y legisladores es crucial para aplicar estas medidas prácticas, asegurando que las respuestas jurídicas ante este fenómeno estén bien fundamentadas, sean eficientes y se adapten a la rápida evolución del panorama digital.

Bibliografía

Ayers, D. (2021). The limits of transactional identity: Whiteness and embodiment in digital facial replacement. *Convergence*, *27*(4), 1018-1037. https://doi.org/10.1177/13548565211027810

Bayer, J., Holznagel, D. B., Katarzyna, L., Adela, P., Josephine, B. S., Szakács, J., & Uszkiewicz, E. (2021). *Disinformation and propaganda: Impact on the functioning of the rule of law and democratic processes in the EU and its Member States: 2021 update*. European Parliament Policy Department for External Relations. https://www.sipotra.it/wp-content/uploads/2021/05/Disinformation-and-propaganda-impact-on-the-functioning-of-the-rule-of-law-and-democratic-processes-in-the-EU-and-its-Member-States-2021-update.pdf

BBC News. (2023, septiembre 23). AI-generated naked child images shock Spanish town of Almendralejo. *BBC News*. https://www.bbc.com/news/world-europe-66877718

Berenzen, P. (2023). *Deepfakes As A Threat To Democracy: Perceptions, Challenges, And Implications Of Deepfake Discourses In Democracies* [Bachelor Thesis, University of Twente]. https://essay.utwente.nl/96206/

Bode, L., Lees, D., & Golding, D. (2021). The Digital Face and Deepfakes on Screen. *Convergence*, *27*(4), 849-854. https://doi.org/10.1177/13548565211034044

Busacca, A., & Monaca, M. A. (2023). Deepfake: Creation, Purpose, Risks. En D. Marino & M. A. Monaca (Eds.), *Innovations and Economic and Social Changes due to Artificial Intelligence: The State of the Art* (pp. 55-68). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-33461-0_6

Casals, M. M. (2023). Las Propuestas de la Unión Europea para regular la responsabilidad civil por los daños causados por sistemas de inteligencia artificial. *InDret*. https://doi.org/10.31009/InDret.2023.i3.02

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, *23*(7), 2072-2098. https://doi.org/10.1177/1461444820925811

Duminica, R., & Ilie, D. M. (2023). Ethical and Legal Aspects of the Development and Use of Robotics and Artificial Intelligence. Protection of Human Rights in the Era of Globalization and Digitisation. *JL & Admin. Sci.*, *19*, 20. https://jolas.ro/wp-content/uploads/2023/06/jolas19a3.pdf

El País México. (2023, octubre 14). *Acusado un universitario de alterar con inteligencia artificial miles de imágenes de alumnas para venderlas como pornografía*. https://elpais.com/mexico/2023-10-14/acusado-un-universitario-de-alterar-con-inteligencia-artificial-miles-de-imagenes-de-alumnas-para-venderlas-como-pornografia.html

Etienne, H. (2021). The future of online trust (and why Deepfake is advancing it). *AI and Ethics*, *1*(4), 553-562. https://doi.org/10.1007/s43681-021-00072-1

Farid, H. (2022). Creating, Using, Misusing, and Detecting Deep Fakes. *Journal of Online Trust and Safety*, *1*(4), Article 4. https://doi.org/10.54501/jots.v1i4.56

Gambín, Á. F., Yazidi, A., Vasilakos, A., Haugerud, H., & Djenouri, Y. (2024). Deepfakes: Current and future trends. *Artificial Intelligence Review*, *57*(3), 64. https://doi.org/10.1007/s10462-023-10679-x

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627-660. https://doi.org/10.5465/annals.2018.0057

Greenough, C. J. (2022). *Make it, Fake it and Get Away with it? The Role of Toxic Masculinity and Threat Perception within Cases, Policies, and Legislation Surrounding Deepfakes* [Master Thesis, University of Auckland]. https://hdl.handle.net/2292/61261

Hameleers, M., van der Meer, T. G. L. A., & Dobber, T. (2024). Distorting the truth versus blatant lies: The effects of different degrees of deception in domestic and foreign political deepfakes. *Computers in Human Behavior*, *152*, 108096. https://doi.org/10.1016/j.chb.2023.108096

Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M. (2023). Deepfake detection using deep learning methods: A systematic and comprehensive

review. *WIREs Data Mining and Knowledge Discovery*, *n/a*(n/a), e1520. https://doi.org/10.1002/widm.1520

Jagatheesaperumal, S. K., Rahouti, M., Ahmad, K., Al-Fuqaha, A., & Guizani, M. (2022). The Duo of Artificial Intelligence and Big Data for Industry 4.0: Applications, Techniques, Challenges, and Future Research Directions. *IEEE Internet of Things Journal*, *9*(15), 12861-12885. https://doi.org/10.1109/JIOT.2021.3139827

Jinana, H. H. A. (2020). *The English torts of defamation and (false) privacy: Analysing the impact of the overlap on defences, interim injunctions and damages* [Ph. D Thesis, Keele University]. https://keele-repository.worktribe.com/output/418965

Karnouskos, S. (2020). Artificial Intelligence in Digital Media: The Era of Deepfakes. *IEEE Transactions on Technology and Society*, *1*(3), 138-147. https://doi.org/10.1109/TTS.2020.3001312

Khuat, T. T., Kedziora, D. J., & Gabrys, B. (2022). *The Roles and Modes of Human Interactions with Automated Machine Learning Systems* (arXiv:2205.04139). arXiv. https://doi.org/10.48550/arXiv.2205.04139

Kirchengast, T. (2020). Deepfakes and image manipulation: Criminalisation and control. *Information & Communications Technology Law*, *29*(3), 308-323. https://doi.org/10.1080/13600834.2020.1794615

Kugler, M. B., & Pace, C. (2021). Deepfake privacy: Attitudes and regulation. *Nw. UL Rev.*, *116*, 611. https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1476&context=nulr

Langa, J. (2021). Deepfakes, real consequences: Crafting legislation to combat threats posed by deepfakes. *BUL Rev.*, *101*, 761. http://www.bu.edu/bulawreview/files/2021/04/LANGA.pdf

Leiser, M. (2022). Bias, journalistic endeavours, and the risks of artificial intelligence. En T. Pihlajarinne & A. Alén-Savikko (Eds.), *Artificial Intelligence and the Media* (pp. 8-32). Edward Elgar Publishing Ltd. https://doi.org/10.4337/9781839109973.00007

Li, M., & Wan, Y. (2023). Norms or fun? The influence of ethical concerns and perceived enjoyment on the regulation of deepfake information.

Internet Research, *33*(5), 1750-1773. https://doi.org/10.1108/INTR-07-2022-0561

Lu, Y., & Ebrahimi, T. (2024). Assessment framework for deepfake detection in real-world situations. *EURASIP Journal on Image and Video Processing*, *2024*(1), 6. https://doi.org/10.1186/s13640-024-00621-8

Lussier, N. (2022). Nonconsensual Deepfakes: Detecting and Regulating This Rising Threat to Privacy. *Idaho L. Rev.*, *58*, 353. https://digitalcommons.law.uidaho.edu/cgi/viewcontent.cgi?article=1252&context=idaho-law-review

Mahashreshty, S. (2023). *Implications of Deepfake Technology on Individual Privacy and Security* [Master Thesis, St. Cloud State University]. https://repository.stcloudstate.edu/msia_etds/142

Mammadzada, I. (2021). *Deepfakes And Freedom Of Expression: European Perspective* [Master's thesis, Tallinn University of Technology]. https://digikogu.taltech.ee/et/Download/ec6ea9ff-bd47-49e4-986d-3a115c00300b/SvavltsingjavljendusvabadusEuroopaperspekt.pdf

Mekkawi, M. H. (2023). The challenges of Digital Evidence usage in Deepfake Crimes Era. *Journal of Law and Emerging Technologies*, *3*(2), Article 2. https://doi.org/10.54873/jolets.v3i2.123

Montasari, R. (2024). Responding to Deepfake Challenges in the United Kingdom: Legal and Technical Insights with Recommendations. En R. Montasari (Ed.), *Cyberspace, Cyberterrorism and the International Security in the Fourth Industrial Revolution: Threats, Assessment and Responses* (pp. 241-258). Springer International Publishing. https://doi.org/10.1007/978-3-031-50454-9_12

Mukta, M. S. H., Ahmad, J., Raiaan, M. A. K., Islam, S., Azam, S., Ali, M. E., & Jonkman, M. (2023). An Investigation of the Effectiveness of Deepfake Models and Tools. *Journal of Sensor and Actuator Networks*, *12*(4), Article 4. https://doi.org/10.3390/jsan12040061

Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A., & Dwivedi, Y. K. (2023). Deepfakes: Deceptions, mitigations, and opportunities.

Journal of Business Research, *154*, 113368. https://doi.org/10.1016/j.jbusres.2022.113368

Naitali, A., Ridouani, M., Salahdine, F., & Kaabouch, N. (2023). Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions. *Computers*, *12*(10), Article 10. https://doi.org/10.3390/computers12100216

Ng, G. W., & Leung, W. C. (2020). Strong Artificial Intelligence and Consciousness. *Journal of Artificial Intelligence and Consciousness*, *07*(01), 63-72. https://doi.org/10.1142/S2705078520300042

Nikolakopoulos, A., Julian Segui, M., Pellicer, A. B., Kefalogiannis, M., Gizelis, C.-A., Marinakis, A., Nestorakis, K., & Varvarigou, T. (2023). BigDaM: Efficient Big Data Management and Interoperability Middleware for Seaports as Critical Infrastructures. *Computers*, *12*(11), Article 11. https://doi.org/10.3390/computers12110218

OCDE. (2019). *Principios de la OCDE sobre Inteligencia Artificial*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

O'Halloran, A. (2021). *The Technical, Legal, and Ethical Landscape of Deepfake Pornography* [Thesis for Bachelor of Science in Computer Science, Brown University]. https://cs.brown.edu/media/filer_public/57/4a/574aeab5-d4dd-4507-b9b6-6c36a04b6ed8/ohalloranamelia.pdf

Okolie, C. (2023). Artificial Intelligence-Altered Videos (Deepfakes), Image-Based Sexual Abuse, and Data Privacy Concerns. *Journal of International Women's Studies*, *25*(2). https://vc.bridgew.edu/jiws/vol25/iss2/11

Pantserev, K. A. (2020). The Malicious Use of AI-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability. En H. Jahankhani, S. Kendzierskyj, N. Chelvachandran, & J. Ibarra (Eds.), *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity* (pp. 37-55). Springer International Publishing. https://doi.org/10.1007/978-3-030-35746-7_3

Pascale, E. (2023). Deeply Dehumanizing, Degrading, and Violating: Deepfake Pornography and the Path to Legal Recourse. *Syracuse L. Rev.*, *73*, 335. https://lawreview.syr.edu/wp-content/uploads/2023/03/Pascale-335-366.pdf

Pavis, M. (2021). Rebalancing our regulatory response to Deepfakes with performers' rights. *Convergence*, *27*(4), 974-998. https://doi.org/10.1177/13548565211033418

Pawelec, M. (2022). Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *Digital Society*, *1*(2), 19. https://doi.org/10.1007/s44206-022-00010-6

Peru21. (2023, septiembre 2). El lado oscuro de la Inteligencia Artificial: La manipulación de voz e imágenes | PERU. *Peru21*. https://peru21.pe/peru/el-lado-oscuro-de-la-inteligencia-artificial-la-manipulacion-de-voz-e-imagenes-noticia/

Pombo, C., Cabrol, M., Alarcón, N. G., & Ávalos, R. S. (2020). fAIr LAC: Adopción ética y responsable de la inteligencia artificial en América Latina y el Caribe. *IDB Publications*. https://doi.org/10.18235/0002169

Popova, M. (2020). Reading out of context: Pornographic deepfakes, celebrity and intimacy. *Porn Studies*, *7*(4), 367-381. https://doi.org/10.1080/23268743.2019.1675090

Rizzica, A. (2021). *Sexually explicit deepfakes: To what extent do legal responses protect the depicted persons?* [Master's Thesis Law and Technology, Tilburg University]. http://arno.uvt.nl/show.cgi?fid=154764

Schick, N. (2020). *Deepfakes: The coming infocalypse*. Hachette UK.

Singh, D. P., & Dhiman, D. B. (2023). *Exploding AI-Generated Deepfakes and Misinformation: A Threat to Global Concern in the 21st Century*. TechRxiv. https://doi.org/10.36227/techrxiv.24715605.v1

Staab, L. C. (2023). *Making Sense of Deepfakes: Epistemic Harms and the EU Policy Response* [Thesis Project, Universiteit Twente]. https://essay.utwente.nl/97795/

Tahraoui, M., Krätzer, C., & Dittmann, J. (2023). Defending Informational Sovereignty by Detecting Deepfakes: Risks and Opportunities of an AI-Based Detector for Deepfake-Based Disinformation and Illegal Activities. En B. Herlo & D. Irrgang (Eds.), *Proceedings of the Weizenbaum Conference 2022: Practicing Sovereignty—Interventions for Open Digital Futures* (pp. 142-161). Weizenbaum Institute for the Networked Society - The German Internet Institute. https://doi.org/10.34669/wi.cp/4.14

Tariq, S., Iftikhar, A., Chaudhary, P., & Khurshid, K. (2023). Is the 'Technological Singularity Scenario' Possible: Can AI Parallel and Surpass All Human Mental Capabilities? *World Futures*, *79*(2), 200-266. https://doi.org/10.1080/02604027.2022.2050879

UNESCO. (2022). *Recomendación sobre la ética de la inteligencia artificial*. https://www.unesco.org/es/articles/recomendacion-sobre-la-etica-de-la-inteligencia-artificial

US Congress. (2022, mayo 24). *S.2559 - 117th Congress (2021-2022): Deepfake Task Force Act* (2021-07-29) [Legislation]. Summary of S.2559 - 117th Congress (2021-2022): Deepfake Task Force Act. https://www.congress.gov/bill/117th-congress/senate-bill/2559/summary/00

Vasist, P. N., & Krishnan, S. (2022). Deepfakes: An Integrative Review of the Literature and an Agenda for Future Research. *Communications of the Association for Information Systems*, *51*(1). https://doi.org/10.17705/1CAIS.05126

Waseem, S., Abu Bakar, S. A. R. S., Ahmed, B. A., Omar, Z., Eisa, T. A. E., & Dalam, M. E. E. (2023). DeepFake on Face and Expression Swap: A Review. *IEEE Access*, *11*, 117865-117906. https://doi.org/10.1109/ACCESS.2023.3324403

Whittaker, L., Mulcahy, R., Letheren, K., Kietzmann, J., & Russell-Bennett, R. (2023). Mapping the deepfake landscape for innovation: A multidisciplinary systematic review and future research agenda. *Technovation*, *125*, 102784. https://doi.org/10.1016/j.technovation.2023.102784