Juan Emilio Suñé Cano
Universidad Camilo Jose Cela
jsune@ucjc.edu 0000-0001-9569-0726

Pablo Fernández Alonso
Universidad Camilo Jose Cela
pablo.falonso@ucjc.edu 0000-0003-0742-882X

■ Recibido / Received
31 de octubre de 2024

■ Aceptado / Acepted
27 de noviembre de 2024

■ Páginas / Pages
De la 41 a la 64

■ ISSN: 1885-365X

El lado oscuro de los avatares generados con inteligencia artificial. El *deepfake* en la pornografía

The dark side of artificial intelligence-generated avatars. Deepfake in pornography

RESUMEN:
El artículo explora el impacto de los deepfakes en la creación de avatares digitales y su uso en la pornografía. Los deepfakes permiten crear alteraciones convincentes de videos, audios e imágenes, lo que plantea serias preocupaciones sobre la privacidad, el consentimiento y la explotación. El estudio analiza cómo los deepfakes pueden crear contenido explícito sin el consentimiento de las personas, generando preocupaciones éticas y legales. Se utiliza una investigación jurídica para analizar las implicaciones legales de los deepfakes y se explica cómo funcionan estas tecnologías, como las redes neuronales generativas antagónicas (GAN). Se discute cómo los deepfakes se utilizan para crear pornografía no consensuada, afectando gravemente la privacidad y la dignidad de las personas. El documento también analiza las leyes actuales y la necesidad de una regulación más estricta para abordar el uso malintencionado de los deepfakes. Se destaca la urgencia de desarrollar marcos legales robustos y la responsabilidad de las plataformas tecnológicas para proteger a las víctimas. Además, se subraya la necesidad de mejores herramientas de detección y de educación pública sobre cómo identificar este tipo de contenido. En conclusión, el estudio resalta la necesidad de un enfoque integral para mitigar los riesgos asociados con el abuso de estas tecnologías y proteger los derechos fundamentales de las personas en la era digital.

PALABRAS CLAVE:
Deepfakes, avatar, sexualización, delitos, pornografía, metaverso.

ABSTRACT:
The article explores the impact of deepfakes on the creation of digital avatars and their use in pornography. Deepfakes make it possible to create convincing alterations of videos, audio and images, raising serious concerns about privacy, consent and exploitation. The study analyzes how deepfakes can create explicit content without people's consent, raising ethical and legal concerns. Legal research is used to analyze the legal implications of deepfakes and explain how these technologies, such as generative adversarial neural networks (GAN), work. It is discussed how deepfakes

are used to create non-consensual pornography, seriously affecting people's privacy and dignity. The document also discusses current laws and the need for stricter regulation to address the malicious use of deepfakes. The urgency of developing robust legal frameworks and the responsibility of technological platforms to protect victims is highlighted. Additionally, the need for better detection tools and public education on how to identify this type of content is highlighted. In conclusion, the study highlights the need for a comprehensive approach to mitigate the risks associated with the abuse of these technologies and protect the fundamental rights of people in the digital age.

KEY WORDS:

Deepfakes, avatar, sexualization, crimes, pornography, metaverse.

1. Introducción a los avatares *deepfake* y la pornografía: Un estudio académico

El rápido avance de la inteligencia artificial (IA) y las tecnologías de aprendizaje automático ha permitido importantes progresos en la creación de medios sintéticos, especialmente los *deepfakes*. Estas manipulaciones generadas por IA pueden crear alteraciones altamente convincentes de vídeos, audios e imágenes, haciendo cada vez más difícil distinguir entre la realidad y la fabricación. En particular, el uso de la tecnología *deepfake* en el contexto de los avatares —representaciones digitales de personas reales— se ha convertido en un tema crítico tanto en el ámbito del entretenimiento como en el discurso ético.

Una de las aplicaciones más controvertidas de la tecnología *deepfake* es su intersección con la pornografía. Los avatares *deepfake* pueden ser utilizados para crear contenido explícito al superponer la imagen de una persona —frecuentemente sin su consentimiento— sobre un cuerpo digital o real. Esta práctica ha generado serias preocupaciones sobre la privacidad, el consentimiento y la explotación, lo que plantea preguntas éticas, sociales y legales sobre la regulación de las tecnologías basadas en IA y su impacto en la vida de las personas.

El surgimiento de avatares *deepfake* en la pornografía ha dado lugar a una serie de debates en contextos académicos, legales y culturales. Por ejemplo, la proliferación de la pornografía *deepfake* no consensuada se ha convertido en un tema urgente, ya que la capacidad de los perpetradores para manipular imágenes de personas —ya sean figuras públicas o ciudadanos «anónimos»— para crear contenido explícito causa un daño psicológico, social y emocional significativo. Mientras que los avatares generados por IA en la industria del videojuego y el entretenimiento suelen promover la libertad personal y la expresión creativa, su aplicación en la pornografía ha suscitado preocupaciones sobre la explotación, el consentimiento y la mercantilización de la imagen de las personas.

Este estudio tiene como objetivo explorar las implicaciones de los avatares *deepfake* en la pornografía, centrándose en las ramificaciones legales, sociales y psicológicas de tal tecnología. Se examinarán las siguientes preguntas: ¿Cómo contribuyen los avatares *deepfake* al auge del contenido explícito no consensuado? ¿Cuáles son las consideraciones legales sobre el uso de la IA para generar pornografía sintética? ¿Cómo afectan tales avances en diversas áreas? Además, se discutirá el esfuerzo de los legisladores, las empresas tecnológicas y los grupos de defensa para abordar los desafíos planteados por la pornografía *deepfake* y las posibles soluciones que se están explorando.

Al analizar las intersecciones de la tecnología *deepfake*, los avatares y la pornografía, este estudio contribuirá a una comprensión más amplia de las consecuencias de estos avances, ofreciendo una visión de los marcos legales que están emergiendo para proteger la privacidad y el consentimiento en un mundo cada vez más digitalizado.

2. Metodología

Siendo el área de conocimiento sobre la que versa este artículo el Derecho, se seguirá para su realización la metodología de la investigación jurídica. No obstante, dado que tal metodología se define como la «disciplina de carácter filosófico que tiene como objeto el estudio de los métodos generales de conocimiento que se pueden utilizar para indagar y resolver problemas vinculados con lo jurídico», se detalla el método utilizado.

Por otro lado, esta tarea requiere ser abordada desde la posición del jurista o científico del Derecho, en el sentido de persona que estudia la interpretación y aplicación del Derecho, por contraposición al profesional del Derecho, ya que, a diferencia de la actividad de estos últimos, no va a versar este trabajo sobre la resolución de un caso concreto.

De esta forma, el método a utilizar será la herramienta que posibilite extraer los conceptos y principios generales cuyo conocimiento es fundamental para comprender las instituciones analizadas de manera completa, superando sus concretas aplicaciones contenidas en las diversas fuentes analizadas.

Así, para realizar este trabajo se emprenderá una investigación documental, consultando y recogiendo información de material impreso y virtualizado.

De esta forma, este trabajo se nutrirá a partir de una revisión de la literatura, entendida esta como un análisis en profundidad de documentos relacionados con el tema que se va a desarrollar.

En cuanto a las etapas de esta investigación documental, siguiendo el modelo de Álvarez Undurraga,[1] se distinguen las siguientes:

- Etapa de planteamiento o aporética. En ella, se selecciona el tema y problema a tratar, los objetivos, el método y la planificación. Así, el resultado de esta etapa se plasma en el presente apartado, en el apartado objetivos y en el cronograma dispuesto en el aula virtual.
- Etapa de erudición o heurística. En esta, se identifican y localizan las fuentes de conocimiento jurídico a utilizar.
- Etapa de construcción o análisis. Así, le corresponde a esta etapa el análisis de datos, su interpretación y su síntesis.
- Etapa de comunicación o formal. Por último, esta etapa supone concluir el trabajo propuesto, estructurando y expresando la información elaborada.

Por último, siendo la fuente principal de información de este trabajo el análisis de situaciones y su respuesta a diversos niveles, el método de investigación fundamental será el inductivo, llegando desde lo particular, esto es, el problema jurídico resuelto por las diferentes instancias implicadas, a lo general, siendo esto las regularidades o patrones que se observan en la aplicación del Derecho a la cuestión analizada.

1/ Álvarez U. G. (2002). Metodología de la investigación jurídica: Hacia una nueva perspectiva (Tesis de licenciatura, Universidad Central de Chile, Facultad de Ciencias Jurídicas y Sociales).

3. Introducción, ¿pero de verdad Taylor Swift apoya a Donald Trump o está en la industria de la pornografía?

A fecha de cierre del presente artículo, quedan días para las elecciones presidenciales de 2024 en los Estados Unidos de América del Norte, una campaña sin duda con muchas vicisitudes a comentar, pero nos centraremos en una: cuando en los meses finales de la campaña tanto el candidato republicano, Donald Trump, como sus seguidores utilizaron la imagen de una persona en concreto para hacer campaña: Taylor Swift. Pero no fotografías reales, sino vídeos e imágenes manipuladas o con avatares virtuales simulando o con inmenso parecido a Taylor Alison Swift (nombre completo de la artista) generadas con IA, los conocidos como *deepfakes* en los que ahora nos adentraremos, como sus propios autores reconocieron posteriormente, en las que la cantante supuestamente muestra su apoyo al candidato republicano. Era una doble falsedad, primero por esos avatares o imágenes manipuladas simulando ser la artista y por difundir un mensaje falso, ya que posteriormente al único debate electoral entre la vicepresidenta Kamala Harris y Donald Trump, la cantante publicó un *post* en Instagram en el que da su apoyo a Harris y afirmaba que votaría por ella en las elecciones de noviembre.

Pero no era la primera vez, ni será la última, en la que la artista sufría la utilización de avatares virtuales o imágenes alteradas con IA sobre su imagen, y lo había padecido de un modo mucho más oscuro aun si cabe, el de la pornografía. En enero de 2024, imágenes sexualmente explícitas de Swift generadas con IA circularon con muchísimo impacto en Twitter (ahora X). Una de las publicaciones (de un usuario verificado) llegó a tener 45 millones de visualizaciones, 24 000 republicaciones y cientos de miles de cuentas guardaron el tuit antes de que se suspendiera la cuenta 17 horas más tarde por, según publicó X, incumplir la prohibición de «publicar imágenes de desnudez no consensuada». El contenido sexual consentido sí está permitido en la red social.

Pero eso fue solo el principio, y no todas las imágenes sexuales generadas de la cantante se retiraron de la red social: las imágenes se difundieron en otras cuentas y en otras plataformas, aparecieron nuevos contenidos y *deepfakes* manipulados de contenido pornográfico, y el término 'Taylor Swift AI' fue viral en varias regiones del mundo.

No en vano, la artista fue nominada Persona del año en 2023 para la revista *Time*, y se la considera la quinta mujer más poderosa del mundo según Forbes. Con 283 millones de seguidores en Instagram, y siendo la primera fortuna femenina en el mundo empresarial de la música tras recientemente desbancar a la artista y empresaria barbadense Robyn Rihanna Fenty, conocida artísticamente como Rihanna, Taylor Swift es ya la cantante más rica del mundo tras haber superado los 1400 millones de dólares de Rihanna. Con tan solo 34 años, el patrimonio de la estadounidense alcanza ya los 1600 millones de dólares (unos 1458 millones de euros), según informó la revista Forbes. Y pese a esta popularidad y posicionamiento empresarial, o tal vez precisamente por ello, padeció este tipo de actos deplorables, haciéndonos una idea de la magnitud y alcance del problema.

Según 404 Media,[2] esas primeras imágenes con avatares sexualizados de la empresaria estadounidense podrían haber surgido en un grupo de Telegram donde los usuarios comparten imágenes explícitas de mujeres generadas con la herramienta que incluye IA Microsoft Designer. Pero no han sido las únicas. Con la aparición de la nueva IA de X, Grok, en la red social se movieron imágenes sexualizadas de Swift en ropa interior y situaciones íntimas. Varias investigaciones señalan que alrededor del 96 % de los *deepfakes* se usan para crear contenido pornográfico de mujeres famosas sin su consentimiento.[3]

4. Mentiras profundas o *deepfakes*, ¿qué son?

Desde la perspectiva de los que suscriben, se podría definir un *deepfake* como un vídeo, foto o grabación de audio que parece real pero que ha sido manipulado con IA. La tecnología subyacente puede reemplazar rostros, manipular expresiones faciales, sintetizar rostros y sintetizar el habla. Los *deepfakes* pueden mostrar a alguien que parece decir o hacer algo que, de hecho, nunca dijo o hizo. Si bien los *deepfakes* tienen aplicaciones benignas y legítimas en áreas como el entretenimiento y el comercio, se utilizan comúnmente para la explotación.

Según un informe reciente de la empresa Deeptrace, gran parte del contenido *deepfake* en línea es pornográfico, y la pornografía *deepfake* victimiza desproporcionadamente a las mujeres. Además, existe preocupación por el posible crecimiento del uso de *deepfakes* para otros fines, en particular la desinformación. Los *deepfakes* podrían usarse para influir en las elecciones o incitar disturbios civiles, o como arma de guerra psicológica. También podrían llevar a que se ignoren las pruebas legítimas de irregularidades y, de manera más general, socavar la confianza pública en el contenido audiovisual.

¿Cómo funciona? Los *deepfakes* se basan en redes neuronales antagónicas o artificiales, de las que hablaremos en profundidad más adelante en este mismo artículo, pero sintetizando son sistemas informáticos modelados libremente sobre el cerebro humano que reconocen patrones en los datos. Desarrollar una fotografía o un vídeo *deepfake* normalmente implica alimentar cientos o miles de imágenes a la red neuronal artificial, 'entrenándola' para identificar y reconstruir patrones, normalmente rostros.

Los *deepfakes* utilizan diferentes tecnologías de IA subyacentes, en particular autocodificadores o redes generativas antagonicas (GAN). Un autocodificador es una red neuronal artificial entrenada para reconstruir la entrada a partir de una representación más simple. Una GAN está formada por dos redes neuronales artificiales que compiten, una que intenta producir una falsificación y la otra que intenta detectarla. Esta competición continúa durante muchos ciclos, lo que da como resultado una representación más plausible de, por ejemplo, rostros en un vídeo. Las GAN suelen producir *deepfakes* más convincentes, pero son más difíciles de usar.

Los investigadores y las empresas de internet han experimentado con varios métodos para detectar *deepfakes*. Estos métodos también suelen utilizar IA para analizar vídeos en busca de artefactos digitales o detalles que los *deepfakes* no consiguen imitar de forma realista, como parpadeos o tics faciales.

2/ https://www.404media.co/microsoft-closes-loophole-that-created-ai-porn-of-taylor-swift/

3/ https://igp.sipa.columbia.edu/news/rise-deepfake-pornography https://www.bbc.com/news/technology-49961089

La gran pregunta que se hará el lector será, sin duda, ¿cuán accesible/fácil de usar es esta tecnología? Cualquier persona con conocimientos informáticos básicos y un ordenador en casa o un móvil de cierta capacidad, puede crear un *deepfake*. Hay aplicaciones informáticas disponibles abiertamente en internet con tutoriales sobre cómo crear vídeos *deepfake*.

No podemos olvidar el caso del Reino de España, donde en septiembre de 2023 se produjo el caso de distribución de imágenes *deepfakes* de alumnas menores de edad en un colegio de Badajoz; los identificados finalmente como autores fueron varios menores de edad que generaron y distribuyeron las imágenes de niñas aparentemente desnudas del mismo colegio en la localidad de Almendralejo, todos ellos menores de 14 años. Fuentes policiales confirmaron que los agentes identificaron a al menos 10 menores, varios de los cuales manejaban la aplicación mediante la cual se creaban las imágenes artificiales, y la mayoría formaban parte del chat en el que se distribuyeron esas fotos.[4]

Desde el punto de vista legal en España, el hecho de que varios de los menores identificados tengan 13 años o menos cierra la puerta a poder exigirles a estos una responsabilidad penal. No obstante, sí que se podría actuar penalmente contra los que tienen ya 14 años, si bien es complicado determinar el tipo de imputación, ya que hasta que no se analicen las imágenes y la intervención de cada implicado, no es posible determinar qué delitos pueden haber cometido los que hayan superado esta edad. El Ministerio público exploraba, en un principio, cuatro tipos. Para que se den los dos primeros, elaboración o distribución de material pornográfico (artículo 189.1b del Código Penal) y tenencia de pornografía infantil (artículo 189.5), el Código Penal exige que las imágenes muestren un acto sexual explícito en el que participa un menor o se vean los órganos sexuales del menor «con fines principalmente sexuales». También podría haber un delito contra la integridad moral de las menores (artículo 173) o uno contra la intimidad (artículo 197) en el caso de que la imagen de la cara de las niñas se haya obtenido invadiendo su intimidad (por ejemplo, mediante fotos tomadas en un espacio privado o sacadas de sus perfiles privados en redes sociales).

Todo lo anterior nos demuestra la facilidad de acceso y manejo de este tipo de herramientas de generación de *deepfake* de carácter pornográfico, y la dificultad de su persecución y penalización.

Sin embargo, para desarrollar un *deepfake* algo realista, estas aplicaciones generalmente todavía requieren cientos o miles de imágenes de entrenamiento de los rostros que se van a intercambiar o manipular, lo que hace que las celebridades y los líderes gubernamentales sean los sujetos más comunes. Los *deepfakes* más convincentes creados con GAN requieren habilidades y recursos técnicos más avanzados. A medida que las tecnologías de redes neuronales artificiales han avanzado rápidamente en paralelo con una computación más potente y abundante, también lo ha hecho la capacidad de producir *deepfakes* realistas.

5. Creación y consecuencias de los *deepfake*

En el año 2017, investigadores de la Universidad de Washington en los Estados Unidos demostraron cómo se podía manipular la tecnología de edición de generación de imágenes dinámicas

4/ https://elpais.com/sociedad/2023-09-21/identificados-10-menores-como-autores-de-los-desnudos-con-inteligencia-artificial-de-almendralejo.html

utilizando un algoritmo de aprendizaje profundo para imitar las expresiones faciales y la voz del presidente Obama, creando vídeos del expresidente que parecen hacer discursos con palabras de entrevistas anteriores.[5] Sin embargo, la generación de imágenes falsas tuvo parte de su origen, que no vamos a considerar aquí, en el uso de imágenes reales de actrices trasplantadas a cuerpos de otras actrices del mercado de la pornografía en redes como Reddit.[6]

La tecnología de IA que hace posible la creación de las *deepfakes* posibilita crear vídeos sofisticados tan realistas que son casi imposibles de distinguir de la realidad. Las mentiras profundas o *deepfakes* son preocupantes precisamente porque permiten la manipulación de la imagen de cualquier persona y ponen en tela de juicio nuestra capacidad de confiar en lo que vemos. Un uso obvio de *deepfakes* sería implicar falsamente a personas en escándalos de las más variadas naturalezas: desde los de carácter político, financieros, sexuales, etc. Incluso si se demuestra que las imágenes incriminatorias son falsas, el daño a la reputación de la víctima puede ser imposible de reparar. Por ejemplo, los políticos podrían recrear viejas imágenes de sí mismos para que pareciera que siempre habían apoyado una narrativa que recientemente se habría hecho popular, actualizando falsamente o recreando sus posiciones políticas en tiempo real; igualmente, sería posible generar de la nada imágenes ficticias de un político en actividades que nunca existieron.

Incluso podrían diseñarse figuras públicas o privadas que son completamente imaginarias, originales, pero no auténticas, es decir, sintéticas o imágenes de síntesis. Mientras tanto, las imágenes de vídeo podrían volverse inútiles como evidencia en los tribunales. Las noticias de difusión audiovisual podrían reducirse a las personas que debaten si los vídeos son auténticos o no, utilizando una IA cada vez más compleja para tratar de detectar este tipo de mentiras profundas que no siempre serán capaces de detectar las manipulaciones mejor elaboradas, como advierten Bunk *et al.* (2017) o Li y Siwei (2018), entre otros. Existen diversos tipos de *deepfakes*, desde aquellos que constituyen *swaps* de rostros (intercambio), los *deepfakes* de audio que imitan una forma de voz, las recreaciones faciales dinámicas completas, o aquellas que sincronizan los labios de la voz falsa y la insertan en un rostro público que reproducirá miméticamente la expresión facial fundida con la oral falsificada, lo que desarrollará modelos miméticos de carácter sintético que recreará elementos de la comunicación no verbal capaces de generar la sensación en el auditorio de que corresponde a la persona real de la que se falsifica esa información gestual característica, entre las más comunes.

Lo que está en juego con la aparición de estas falsificaciones de vídeo profundas es la estructura social subyacente en la que la mayor parte de la sociedad, en un momento dado, está de acuerdo en que existe alguna forma de verdad mutuamente aceptada y ampliamente difundida y las realidades sociales que se basan en esta confianza. Tras la desinformación masiva, incluso las figuras públicas honestas podrían ser fácilmente ignoradas o desacreditadas. Las organizaciones tradicionales que han apoyado y permitido el consenso social y político, el gobierno y la prensa, ya no serán suficientemente aptas para el propósito que habrían venido desarrollando en el entorno no digital. Como señala Frankfurt (2007), las ideas de verdad y facticidad son indispensables para dotar de plena sustantividad al ejercicio de la racionalidad.

5/ https://www.bbc.com/news/av/technology-40598465

6/ https://www.bbc.com/news/av/technology-40598465

Algunas personas cuestionan los hechos en torno a eventos que sin duda sucedieron, tales como el Holocausto, el aterrizaje del hombre en la Luna o los atentados del 11 de septiembre de 2001 en los Estados Unidos o que la Tierra sea redonda, a pesar de las pruebas de toda clase existentes. Si las mentiras profundas logran que las personas crean que no pueden confiar en las imágenes, los problemas de la desinformación y las teorías de conspiración podrían empeorar significativamente, como señalan Jolley y Douglas (2017). El pensamiento conspirativo se caracteriza, como señalan Bauer, Bradley y Bangerter (2013), por la incapacidad de asignar a los hechos adversos un determinante causal, lo que implica un modo «casi religioso» de pensar en los procesos. Si bien es cierto que la tecnología con la que se elaboran los *deepfakes* no es, por el momento, lo suficientemente sofisticada como para simular eventos o conflictos históricos a gran escala. Preocupa que la duda planteada por uno o varios *deepfakes* convincentes y bien difundidos a escala internacional que afectan a los nodos apropiados de difusión social —es decir, los más densamente conectados— puedan alterar nuestra confianza en el audio y el vídeo de forma permanente.

También son cada vez más fáciles y baratos de crear los programas para la elaboración y desarrollo de este tipo vídeos, lo que significa que pronto será posible que cualquier persona con un ordenador personal y la capacidad apropiada de procesamiento —lo que no excluye procesamiento en red por grupos para disponer de mayor capacidad de cálculo en menos tiempo— y el *software* adecuado, dispongan de los medios necesarios para crearlas y difundirlas de forma acelerada y eficiente en cualquier parte del mundo.

Es cierto que existe una acción cada vez más acentuada para intentar detectar este tipo de mentiras profundas; así, se puede observar el esfuerzo del proyecto liderado por la Agencia de Proyectos de Investigación Avanzados de Defensa (DARPA), que es una agencia del Departamento de Defensa de Estados Unidos responsable del desarrollo de nuevas tecnologías para uso militar, denominado Media Forensics (MediFor). El programa Media Forensics está desarrollando herramientas capaces de identificar cuándo los vídeos y las fotos han sido alterados significativamente de su estado original para cambiar su contenido; este programa viene desarrollándose desde el año 2015, cuando se evidenció como problema de seguridad nacional en los Estados Unidos.

Señalaba Richard Feynman[7] que la ciencia es una larga historia de aprender la manera de no engañarnos, quizá por ello necesitaremos desarrollar nuevas formas de consenso, nuevas maneras de ponerse de acuerdo sobre situaciones sociales basadas en formas alternativas de confianza. Un enfoque prometedor, pero no exento de limitaciones, advirtamos, podría ser descentralizar la confianza, de modo que ya no necesitemos algunas instituciones clásicas para garantizar si la información es genuina, papel que ha venido siendo tradicionalmente desempeñado por la prensa o la televisión y la radio en sus diversas dimensiones, tanto en formatos materiales como inmateriales. Y, en cambio, se puede proponer confiar en redes de personas u organizaciones con buena reputación, pensemos como modelo en la Wikipedia, una enciclopedia virtual tan rigurosa como la enciclopedia británica. Una forma de hacer esto podría ser mediante el uso del *blockchain*, la tecnología que impulsa Bitcoin y otras criptomonedas. *Blockchain* funciona creando un libro de contabilidad público almacenado

7/ https://www.hindustantimes.com/more-lifestyle/infectious-enthusiastic-relevant-a-physicist-s-take-on-richard-feynman/story-jD3jualkHCX5WuXcCp0ndL.html

en varias computadoras de todo el mundo a la vez y a prueba de manipulaciones mediante la criptografía. Sus algoritmos permiten a las computadoras acordar la validez de cualquier cambio en el libro de contabilidad, lo que hace que sea mucho más difícil registrar información falsa. De esta manera, la confianza se distribuye entre todas las computadoras que pueden escrutarse mutuamente, aumentando la responsabilidad y haciendo posible, en hipótesis, construir mecanismos de verificación y contraste de fuentes y hechos de fiabilidad contrastada.

6. Las redes neuronales generativas antagónicas

La IA, y en particular las redes generativas antagónicas (GAN), están siendo progresivamente capaces de identificar cosas u objetos con gran precisión; es factible mostrarles diez millones de fotografías y estas, las GAN, podrán identificar con asombrosa precisión en cuáles de ellas aparece, por ejemplo, una persona montando en bicicleta circulando por una calle. El problema es que para crear algo completamente nuevo hace falta imaginación, circunstancia que hasta ahora no era posible en virtud de los modelos y procedimientos disponibles en IA.

El enfoque GAN emplea dos redes neuronales —modelos matemáticos simplificados del cerebro— que se enfrentan entre sí; es decir, la «confrontación entre redes» es la esencia del aprendizaje autónomo producto de esa confrontación. Ambas redes están entrenadas con el mismo conjunto de datos. Una red conocida como la generativa tiene la tarea de crear variaciones en las imágenes que ya ha visto, tal vez una imagen de una bicicleta con una rueda de más. La segunda, conocida como el discriminador, debe identificar si la imagen que está viendo pertenece al conjunto de entrenamiento original o, si, por el contrario, es una imagen falsa producida por la red generativa. A la red discriminadora básicamente se le formula la siguiente cuestión: ¿Es probable que una bicicleta con tres ruedas sea real? Con el tiempo, a la red generativa se le da tan bien producir imágenes que a su pareja discriminadora le resulta imposible detectar la falsificación. En resumen: la red generativa aprende a reconocer y posteriormente a crear imágenes de bicicletas de aspecto realista.

Esta tecnología se ha convertido en uno de los avances más prometedores de la IA en la última década, capaz de ayudar a las máquinas a producir resultados que engañan incluso a los humanos expertos. Las GAN se usaron para crear sonidos e imágenes hiperrealistas. En un convincente ejemplo, los investigadores del fabricante de chips gráficos Nvidia entrenaron a una GAN con fotografías de personas famosas para que el sistema fuera capaz de crear cientos de rostros creíbles de personas que no existen. Otro grupo de investigación consiguió generar pinturas falsas parecidas a las obras de Van Gogh. Si se les fuerza aún más, las GAN pueden reinterpretar las imágenes de diferentes maneras: pueden hacer que una carretera soleada parezca nevada o convertir caballos en cebras.

Los resultados no siempre son perfectos, las GAN pueden crear bicicletas con dos tipos de manillar o tres sillines, por ejemplo, o caras con cejas en el lugar incorrecto del rostro humano. Pero debido a que las imágenes y los sonidos son, por lo general, extraordinariamente realistas, algunos expertos creen que hay una lógica detrás de cómo las GAN comienzan a comprender la estructura subyacente del mundo que ven y oyen. Otros, en cambio, a quienes nos adherimos, piensan que tanto los algoritmos como las redes neuronales que emplean estas tecnologías no son en absoluto conscientes de su propia existencia, carecen de inteligencia y de autopercepción o conocimiento de sí mismas, características esenciales de la

inteligencia humana; son emulaciones y simulaciones rudimentarias, pero sencillamente no saben lo que hacen, circunstancia que sí comprenden quienes las diseñan. El potencial de las GAN es muy grande porque pueden aprender a imitar cualquier distribución de datos. Es decir, se puede enseñar a las GAN a crear mundos inquietantemente similares a los nuestros en cualquier dominio: imágenes, música, habla, prosa, etc. Los tipos de redes GAN evolucionan de forma constante, por ejemplo, las redes recycle-GAN son capaces de traducir el contenido de un dominio a otro, preservando el estilo nativo del primer dominio, es decir, si los contenidos del discurso de un participante se transfieren a otro, se transfieren con el estilo propio del primero al segundo en un proceso de aprendizaje profundo automatizado sin supervisión.

Si la fábula de Pedro y el lobo tuviera lugar hoy, el problema no sería que los campesinos no creyeran a Pedro, sino que hasta se pondrían de parte del lobo. Y es que la moraleja sobre las consecuencias de no ser honesto empieza a desdibujarse a medida que la IA generativa se vuelve más capaz de fabricar contenidos falsos hiperrealistas o *deepfakes*, y estos, a su vez, se convierten en arma arrojadiza para generar un clima de escepticismo generalizado.

«Durante el año pasado, esta nueva tecnología se utilizó en al menos 16 países para sembrar dudas, difamar a los oponentes o influir en el debate público», advierte la edición de este año del informe *Libertad en la Red* elaborado por la organización sin ánimo de lucro Freedom House. Centrándose solo en Europa, es fácil encontrar casos de *deepfakes* enturbiando cuestiones políticas y sociales: Francia, Alemania, Reino Unido y, por supuesto, Estados Unidos, figuran en la lista. «En Francia, una imagen manipulada por IA de un anciano golpeado por la policía circuló por Internet durante las protestas de marzo de 2023, a menudo junto a comentarios que menospreciaban al presidente Emmanuel Macron», señala el informe.

Lo grave no es solo que la tecnología permita disfrazar mentiras de verdades con fines perversos, como sucede habitualmente con las pornovenganzas generadas con IA. Además, en este nuevo mundo en el que todo puede ponerse en duda, importantes sucesos reales podrían fácilmente acabar desmentidos equivocadamente si alguna de las partes interesadas los tilda de bulos generados con IA. Y ¿sabe quién es el único que sale ganando cuando toda verdad parece mentira y toda mentira parece verdad? El lobo.

Es como si el animal lograra convencer a los campesinos de que no se ha comido al rebaño de Pedro argumentando que los restos de ovejas muertas son parte de una farsa orquestada por el chaval para perjudicarle. El fenómeno se conoce como «el beneficio del mentiroso», en el que: «la cautela generalizada ante las falsedades sobre un tema determinado puede enturbiar las aguas hasta el punto de que la gente no crea en las afirmaciones verdaderas», explica el informe, y añade: «Por ejemplo, políticos han calificado informes fiables como falsificaciones habilitadas por la IA, o han difundido contenido manipulado para sembrar dudas sobre contenido genuino muy similar».

En realidad, esto es algo que ya sucedía sin ayuda de la IA. Donald Trump no necesitó *deepfakes* para convencer a una turba desquiciada de que las elecciones de 2020 habían sido manipuladas, a pesar de que no tenía ninguna prueba que lo justificara. El problema es que «las herramientas basadas en IA que pueden generar texto, audio e imágenes se han vuelto más sofisticadas, accesibles y fáciles de usar rápidamente, lo que ha provocado una preocupante escalada de estas tácticas de desinformación», señala el texto.

«Los deepfakes sí representan una amenaza para la política, pero en este momento la más tangible es el hecho de acusar a los deepfakes para hacer que lo real parezca falso»,

alertaba ya en 2019 el investigador de Deeptrace Lab Henry Ajder, a raíz de la publicación de su propio informe sobre el tema. Su trabajo exponía el suceso que ocurrió en Gabón a finales de 2018, cuando la población empezó a dudar de si su entonces presidente, Ali Bongo, seguía vivo. Tras meses sin aparecer en público, su tradicional vídeo para felicitar el año nuevo a los ciudadanos no hizo más que avivar los rumores sobre su supuesta muerte y encubrimiento por parte del gobierno ante quienes estaban convencidos de que era falso.

Bongo no solo sigue vivito y coleando, sino que ocupó el cargo de presidente hasta finales de agosto de 2023. Lamentablemente, ese tipo de situaciones se producen con cada vez más frecuencia. Entre los casos más recientes identificados por Freedom House está el que salpicó al funcionario indio Palanivel Thiagarajan en abril de este año tras la filtración de unas grabaciones en las que se le escuchaba menospreciar a sus compañeros. Thiagarajan denunció que los audios habían sido generados por máquinas, a pesar de que varios investigadores independientes determinaron que al menos uno era auténtico.

Los *deepfakes* reales ya llevan tiempo socavando la sociedad que conocemos por sí solos. Su capacidad de suplantar la identidad de alguien en fotografía, audio o vídeo está siendo explotada por los ciberdelincuentes para perpetrar fraudes y robos millonarios. Por no hablar de los cada vez más graves y frecuentes delitos de *deepfakes* pornográficos generados con IA generativa. Así, una investigación independiente publicada por Wired, advierte de que la publicación de este tipo de vídeos falsos y no consentidos, y que perjudican desproporcionadamente a las mujeres, se ha disparado. Recoge el citado medio:

> Se han vuelto omnipresentes. Según el investigador, que solicitó el anonimato para evitar ser atacado en línea, se han subido al menos 244 625 vídeos a los 35 principales sitios web creados exclusiva o parcialmente para albergar deepfakes pornográficos en los últimos siete años. Durante los primeros nueve meses de este año, se subieron 113 000 vídeos a los sitios web, un aumento del 54 % respecto a los 73 000 subidos en todo 2022. Para fines de este año, el análisis prevé que se habrán producido más vídeos en 2023 que el número total de cada dos años combinados.

Para más inri, el beneficio del mentiroso supone otra vuelta de tuerca a toda esta problemática. Por ejemplo, en el caso de los robos en los que la voz de algún responsable corporativo fue replicada y utilizada para convencer a sus trabajadores para que realizaran transferencias indebidas, cabría la posibilidad de que dicho fraude hubiera sido perpetrado de verdad por el ejecutivo en cuestión, quien intentaría librarse de la culpa alegando una suplantación de identidad mediante inteligencia artificial que, en realidad, nunca habría tenido lugar.

El fondo de esta crisis, especialmente acusada en el mundo de la política y los negocios, es que «es un arma más para los poderosos, que ahora pueden responder con: 'Es un deepfake', ante cualquier cosa que pueda demostrar corrupción y abusos contra los derechos humanos», advirtió también en 2019 el experto en abusos contra los derechos humanos de la organización Witness, Sam Gregory, a raíz de la publicación del informe de Deeptrace Lab. En esa misma pieza, el experto en desinformación y director de la organización sin ánimo de lucro Thoughtful Technology Project, Aviv Ovadya, aseguraba que esta era su mayor preocupación en torno a los *deepfakes* y añadía: «Es bueno cuestionar la evidencia. Pero lo que [los creadores de desinformación] realmente quieren no es que cuestionemos más, sino que lo cuestionemos todo».

Y es que, además de librarse de posibles acusaciones, que la sociedad viva en un clima de escepticismo generalizado es lo que da beneficios a los mentirosos. «Los políticos no solo

señalan su propia inocencia, también critican a sus rivales y a los medios de comunicación, lo que incita a sus partidarios a unirse contra la oposición», explica otro estudio del Instituto Tecnológico de Georgia publicado en 2021.

En este contexto, no es de extrañar que Freedom House[8] concluya que «la libertad en Internet a nivel mundial ha empeorado por decimotercer año consecutivo». Junto al auge de los *deepfakes* y su impulso a beneficio del mentiroso, la organización señala otros ataques más clásicos a los derechos humanos perpetrados en internet en los que «la IA no ha desplazado por completo a los métodos más antiguos de control de la información».

El informe detalla:

> Un récord de 41 gobiernos bloquearon sitios web con contenido que debería estar protegido según los estándares de libertad de expresión dentro del derecho internacional de derechos humanos. Incluso en entornos más democráticos, incluidos Estados Unidos y Europa, los gobiernos consideraron o impusieron restricciones al acceso a destacados sitios web y plataformas de redes sociales, un enfoque improductivo ante las preocupaciones sobre la interferencia extranjera, la desinformación y la seguridad en línea.

La conclusión es que, al mismo tiempo que la censura, los abusos y la desinformación siguen aumentando mediante las tácticas a las que nos ha tocado acostumbrarnos a la fuerza, la IA generativa está convirtiéndose en otro erosionador de la democracia y la sociedad, ya sea cuando se utiliza para fabricar mentiras con fines perversos o simplemente cuando se usa como argumento para cuestionar cualquier escándalo real que el público sí debería conocer.

¿La solución? Tristemente, muy complicada. «Nuestras arquitecturas legales y políticas no están diseñadas de manera óptima para responder», lamentaban ya en 2018 dos expertos en Derecho en la investigación en la que acuñaron el término del beneficio del mentiroso. A pesar de los crecientes esfuerzos por regular la IA que estamos viendo últimamente, el hecho de que los problemas que acarrea no hayan hecho más que multiplicarse con los años augura un futuro sombrío, tanto para Pedro como para todos nosotros. Los lobos deben estar relamiéndose…

7. Pornografía y derechos humanos, un mundo que ya era oscuro antes de la inteligencia artificial. El antecedente de Pornhub

Pornhub es un sitio web de pornografía en internet con sede en Montreal, Canadá. Es uno de los varios sitios web de transmisión de vídeos pornográficos propiedad de MindGeek. A junio de 2020, Pornhub era el décimo sitio web con más tráfico del mundo y el tercer sitio web para adultos con más tráfico solo después de XVideos y XNXX. En 2024 era el primer puesto pornográfico del mundo y el vigésimo lugar más visitado de la red a escala global.

Pornhub comenzó en Montreal como sitio de fotografías profesionales y *amateur* en 2007, y hoy en día cuenta con oficinas y servidores en San Francisco, Houston, Nueva Orleans, Londres y Limasol (Chipre).

8/ https://freedomhouse.org/es/article/nuevo-informe-la-manipulacion-de-las-elecciones-y-el-conflicto-armado-marcaron-el-deterioro

En marzo de 2010, MindGeek (conocida entonces como Manwin) compró la compañía, que posee muchos otros sitios web pornográficos. El sitio está disponible internacionalmente, pero ha sido bloqueado por algunos países como Filipinas, Pakistán, China continental e India. Ofrece pornografía de realidad virtual, entre otros productos, y organiza los premios Pornhub anualmente.

El día 8 de marzo de 2022, con motivo del día internacional de la mujer, Pornhub, la mayor página web de contenido pornográfico del mundo, cambiaba su logotipo tiñéndolo de morado y acompañándolo del mensaje «Happy International Women's Day» (Feliz Día Internacional de la Mujer), dejando así claro que el cambio de color no es casual.

Más allá de que Pornhub se haya convertido en la mayor referencia del contenido pornográfico del mundo —con lo que eso implica—, a finales de 2020 la empresa tuvo que borrar en torno a 10 millones de vídeos de su plataforma debido a que estos no estaban verificados. Se podían encontrar en la web vídeos personales o caseros subidos en forma de venganza y sin consentimiento, y muchísimos vídeos *deepfakes* de celebridades de mayor o menor grado de impacto social global o local. Además, se conoció el caso de una persona que había sido violada siendo menor de edad cuyos vídeos estaban subidos en Pornhub.

La compañía ya había sido criticada por respuestas lentas o inadecuadas a algunos de estos incidentes, incluido el alojamiento del canal de alto perfil Girls Do Porn, que se cerró en 2019 luego de una demanda y cargos de tráfico sexual.

Todo empezó con la publicación de una columna de opinión en el periódico *The New York Times*, centrada en las experiencias de víctimas de abuso sexual cuyos vídeos han terminado en Pornhub, y de cómo las víctimas tenían que lidiar con ello, ante un gigante con más visitas que Netflix o Amazon y que no hacía nada por combatirlo.

La columna, en la que Nicolás Kristof acusaba directamente a la plataforma de estar infestada de vídeos de violaciones, monetiza violaciones infantiles, pornografía de venganza, vídeos de cámaras espía de mujeres duchándose, contenido racista y misógino, e imágenes de mujeres asfixiadas en bolsas de plástico. Una búsqueda de «chicas menores de 18 años» (sin espacio) o «14 años» conduce en cada caso a más de 100 000 vídeos. La mayoría no son de niños agredidos, pero muchos sí lo son.

Kristof hacía así referencia a casos como el de una niña de 15 años que desapareció en Florida y su madre la encontró después en Pornhub en 58 vídeos sexuales. U otro caso en California, en el que las agresiones sexuales a una niña de 14 años se publicaron en Pornhub y no fueron denunciadas a las autoridades por la empresa, sino por un compañero de clase que vio los vídeos. «En cada caso, los delincuentes fueron arrestados por las agresiones, pero Pornhub eludió la responsabilidad de compartir los vídeos y sacar provecho de ellos», recordaba Kristof.

Ante el escándalo que provocó la columna de *The New York Times*, la empresa tuvo que hacer un borrado masivo de vídeos al no poder verificar el origen de estos. Una medida con la que reconocía que no había puesto límites para frenar los vídeos de explotación sexual.

Esta columna también motivó el inicio de sendas investigaciones de parte de dos proveedores de pagos en la plataforma de Pornhub, Visa y MasterCard, después de las cuales ambas compañías decidieron retirar el soporte de esta página web.

Como resultado, las tarjetas de crédito y débito de Visa y MasterCard ya no pueden utilizarse para pagar en Pornhub, incluyendo Pornhub Premium, el servicio de suscripción que permite acceder a medio millón de vídeos exclusivos, además de disfrutar de contenido en alta resolución.

Esto no fue del todo inesperado. Tanto Visa como MasterCard tienen una relación algo complicada con la industria del entretenimiento adulto, y también bloquean los pagos en otras plataformas similares. Sin embargo, las consecuencias que esta decisión puede tener para la industria son mucho mayores, hasta el punto de que el temor a una crisis se ha extendido por internet.

En respuesta, tanto al artículo de *The New York Times* como a la decisión de Visa y MasterCard, en diciembre de 2020 Pornhub opta por tomar varias decisiones, incluyendo un cambio en los términos de uso que prohíbe el contenido que no ha sido verificado.

Concretamente, desde entonces los únicos vídeos que se pueden subir a Pornhub son los de los socios de la compañía y los miembros del programa de modelos. En otras palabras, es contenido que está asociado con una persona o una empresa concreta, y no puede ser subido de manera anónima o secreta. Además, todos los vídeos que no cumplen esas condiciones han sido eliminados de la plataforma.

El impacto que la nueva política ha tenido en Pornhub fue relativo. La semana anterior a la decisión, la página presumía de contar con 13,5 millones de vídeos; el día después de aplicar el borrado masivo, esa cifra cayó hasta los 2,9 millones de vídeos. Es una cifra cambiante, que sube y baja constantemente, y en otro momento del día estuvo en los 5,6 millones de vídeos; a día de hoy, si bien ha visto repercutida su afluencia, sigue teniendo la misma posición global, y los estudios recogen que el consumidor hace uso de diferentes páginas, habiendo migrado a otras para contenido *amateur* o sin censura, como X.Hamster y diversas paginas redireccionales muy difíciles de rastrear y exigir responsabilidad.

Lo cierto es que la pornografía *online* se ha convertido en uno de los elementos donde la regulación en internet entrará más pronto que tarde, como así demuestran algunas de las iniciativas que se han visto ya en Francia y que están sobre la mesa en Reino Unido.

En el caso del país galo, el 30 de julio de 2020 se aprobaba la ley para obligar a los usuarios a confirmar su edad si quieren acceder a plataformas *online* con contenido pornográfico en un intento de frenar el consumo de estos vídeos por parte de los menores. Aunque no fue hasta finales del año pasado cuando llegó el fin del plazo para identificar a los usuarios, una obligación que si no se cumplía haría cerrar a Pornhub, Xvideos, Xhamster, Xnxx o Tukif en el país.

Reino Unido, por su parte, quiere seguir el mismo camino. Con la Online Safety Bill, el Gobierno —entre otras medidas— obligará a los usuarios de páginas para adultos a identificarse como mayores de 18 años, una medida para que los menores no puedan acceder a este tipo de contenido. Y España ya se ha planteado una identificación específica para el acceso para páginas pornográficas.

Por su parte, Pornhub presume de que ha impuesto medidas que ni siquiera plataformas como TikTok, Facebook o Instagram tienen, y afirma que se ha convertido en objetivo de críticas no por sus políticas, sino por ser una plataforma de contenido adulto.

8. El *deepfake* en el mundo de la pornografía

Como hemos visto, la legalidad de la pornografía *deepfake* es compleja y varía significativamente según la jurisdicción. Por ejemplo, no existe ninguna ley federal en los Estados Unidos que aborde actualmente el tema, ni tampoco específicamente en la Unión Europea. Sin embargo, varios estados de los Estados Unidos de América y varios estados de la Unión Europea,

como el mencionado de España, han prohibido la creación o distribución de *deepfakes* bajo ciertas condiciones, como cuando se utilizan para crear pornografía no consensuada, busque generar influencia electoral o viole derechos de propiedad intelectual.

Como venimos viendo, la IA y la tecnología *deepfake* han transformado significativamente el ámbito digital, marcando el comienzo de una nueva era en la que ver ya no es creer. Los *deepfakes*, falsificaciones muy realistas creadas con IA, desafían nuestra capacidad de discernir lo que es real de lo que es fabricado.

El estatus legal de los *deepfakes* es un tema en evolución. Actualmente, no existe una legislación integral en ninguna región ni país del mundo que aborde directamente la creación y distribución de *deepfakes*. La legalidad de estas falsificaciones generadas por IA varía de un lugar a otro, y algunos estados imponen restricciones a su creación y distribución, particularmente cuando se utilizan con fines dañinos como la pornografía no consentida o para influir en las elecciones.

Los *deepfakes* operan en una zona legal gris. No son inherentemente ilegales; sin embargo, pueden volverse ilegales si infringen los derechos de propiedad intelectual, violan los derechos personales mediante la creación de pornografía no consensuada, difunden información errónea o representan una amenaza para la seguridad nacional. La especificidad de las leyes relativas a los *deepfakes* depende en gran medida de la jurisdicción. Esta fragmentación a nivel estatal, y en estados complejos (estados federales, *länders*, comunidades autónomas o equivalentes con competencias en la materia), apunta a un enfoque irregular de la regulación, que puede resultar difícil de manejar.

Una preocupación importante es que más del 90 % del contenido *deepfake* está asociado con material pornográfico, a menudo creado sin consentimiento.[9] A medida que avance la tecnología *deepfake*, es probable que se utilice de otras formas ilegales, como la extorsión y el acoso digital.

La intersección de los *deepfakes* con las leyes de derechos de autor y uso legítimo complica aún más su estatus legal. A medida que los *deepfakes* se vuelven más sofisticados y generalizados, se constata cada vez más evidente la urgencia de una legislación integral para regular esta tecnología en rápida evolución. La falta de leyes uniformes plantea un desafío no solo para los profesionales del Derecho y los legisladores, sino también para las personas afectadas por el uso malicioso de *deepfakes*.

Se debe aclarar que ver *deepfakes* no es ilegal en sí mismo, excepto en los casos en que el contenido incluya material ilegal, como pornografía infantil, y aun aquí en ciertas legislaciones esto sería legalmente debatible, como en el caso del Hentai en Japón, con escenas (cómic o vídeo) de índole sexual con avatares de apariencia claramente de menores de edad, o el caso judicializado de la prohibición de la exhibición temporal en España de la película *A Serbian Movie*, por representar gráficamente la violación de un bebé de corta edad.

La distinción aquí es fundamental: si bien el consumo de contenido *deepfake* no suele generar consecuencias legales para el espectador, la producción y difusión de dicho contenido sin el consentimiento de los sujetos representados puede tener consecuencias legales; así, a medida que los *deepfakes* atraen cada vez más atención, un número creciente de estados han promulgado leyes para regularlos en el contexto de la pornografía no consensuada.

9/ https://www.uoc.edu/es/news/2023/265-deepfakes-pornograficos-cuando-IA-desnuda-tu-intimidad-vulnera-tus-derechos

Países de todo el mundo están lidiando con los desafíos legales que plantean los *deepfakes*. La legislación varía ampliamente y refleja diferentes preocupaciones culturales, legales y políticas.

China se ha posicionado con una fortísima intervención estatal de las redes, ha desarrollado una regulación de *deepfakes* con una legislación que exige el consentimiento del usuario para la producción de *deepfakes* y exige que el contenido generado mediante IA se marque como tal. Así, el 10 de enero de 2023 entraron en vigor unas disposiciones aplicables a proveedores y usuarios de tecnología *deepfake* que regulan procedimientos que abarcan todo el ciclo de vida de esta tecnología, desde su creación hasta su distribución.

Singapur ha adoptado un enfoque diferente al implementar la Ley de Protección contra la Manipulación y Falsedades en Línea, que apunta a declaraciones de hechos falsas en internet. Si bien no está dirigida específicamente a los *deepfakes*, esta ley se les puede aplicar, lo que refleja el compromiso de Singapur de combatir la desinformación.

En el Reino Unido, compartir pornografía *deepfake* se ha declarado ilegal en virtud de la Ley de Seguridad en Línea.

India ha emitido un aviso a las plataformas de redes sociales para que se protejan contra los *deepfakes* que violan las normas tecnológicas del país (aunque no son ilegales *per se*).

Corea del Sur ha adoptado medidas drásticas mediante la promulgación de una ley en 2020 que prohíbe la distribución de *deepfakes* que puedan causar perjuicio al interés público, con sanciones que incluyen penas de hasta cinco años de prisión o multas de hasta 43 000 dólares estadounidenses.

En el plano internacional se deben resaltar los avances regulatorios en esta materia, especialmente a nivel de la Unión Europea, con la aprobación del Reglamento de Inteligencia Artificial, que aborda la tecnología *deepfake* y establecerá una serie de pautas para su uso.

Además, el reciente Reglamento Europeo de Servicios Digitales también hace referencia a la tecnología *deepfake* al imponer la obligación a los motores de búsqueda en línea y a las plataformas en línea de muy gran tamaño de etiquetar los *deepfakes* como tal, de conformidad con el principio de transparencia.

En este contexto, el Servicio de Estudios del Parlamento Europeo emitió el informe *Tackling deepfakes in European policy*, referenciado en la exposición de motivos de esta Proposición de Ley. Este informe examina los riesgos asociados con los *deepfakes*, como el robo de identidad, la intimidación y la afectación de derechos fundamentales como la igualdad y la no discriminación, proponiendo medidas para abordar estos riesgos en diversas dimensiones (entre otras, la tecnológica, la creativa o la de difusión).

A medida que la tecnología *deepfake* continúa evolucionando, es posible que la comunidad internacional deba considerar estrategias más unificadas para abordar sus implicaciones generalizadas.

La aparición de imágenes de desnudos generadas por IA ha introducido desafíos complejos. La cuestión central es si una imagen de desnudo debe ser «real» para que una víctima busque reparación legal. La IA ahora puede crear o manipular imágenes para producir «desnudos» convincentes de individuos reales que nunca dieron su consentimiento ni participaron en la creación del contenido. Aunque estas imágenes no son auténticas, su uso potencial para la pornografía de venganza es una preocupación muy real.

Actualmente, al menos, sí parecen darse posiciones comunes a escala global, y el tratamiento de la pornografía infantil en el contexto de los *deepfakes* es inequívoco: es ilegal, lo

que refleja la postura de que cualquier representación sexualizada de una menor identificable causa daño, independientemente de si la imagen es real o virtual.

Las víctimas de pornografía *deepfake* también pueden explorar en la inmensa mayoría de los países vías legales como derechos de propiedad intelectual, invasión de la privacidad o demandas por difamación, según las características específicas de su caso.

9. *Deepfakes* y avatares: La nueva frontera de la identidad en el metaverso

El metaverso, una evolución digital que promete transformar la relación humana en entornos virtuales, está proliferando rápidamente, impulsado por la IA y su habilidad para crear avatares personalizables y realistas. Sin embargo, no se trata solo de otro videojuego o plataforma de redes sociales; más bien, el metaverso se perfila como un espacio interactivo donde las actividades van desde la construcción de relaciones, a la formación de comunidades, los negocios y la identidad. Fundamentalmente, en estos espacios se están transformando las experiencias virtuales en mundos extremadamente tecnológicos personalizados, y los avatares humanos en entidades adaptativas y escalables.

El metaverso es un espacio compartido que se extiende a través de plataformas y ofrece experiencias inmersivas en 3D y realidad virtual; los usuarios pueden interactuar en espacios que reflejan el mundo real o son completamente de fantasía. La IA es un ingrediente imperativo en la creación del metaverso, que permite que los personajes y los entornos se generen autónomamente y respondan a entornos compartidos. A través de la IA, el metaverso se convierte en un entorno que también 'aprende' acerca de sus usuarios: detecta patrones de comportamiento, reconoce preferencias y, en general, hace que cada experiencia dentro del metaverso sea única para cada individuo.

Por ejemplo, la IA utiliza algoritmos para determinar cómo un usuario se desplaza dentro del metaverso, a qué tipo de experiencias dedica más tiempo, qué tipo de interacciones emplea y, en base a ello, puede ofrecer recomendaciones de experiencias o cambiar la apariencia del metaverso para que este sea más atractivo y personalizado. De esta forma, el metaverso se vuelve más interactivo con el tiempo y sensible a las demandas de sus usuarios.

Otra característica clave del metaverso es la posibilidad de crear avatares, es decir, representaciones virtuales de un individuo que pueden reflejar su verdadera identidad o proyectar una completamente distinta. La IA ha transformado completamente el proceso de creación de avatares. Anteriormente, era difícil crear un avatar detallado, ya que esto requeriría un gran esfuerzo y tiempo; sin embargo, ahora es posible crear avatares hiperrealistas que utilizan métodos de aprendizaje profundo para imitar los rasgos físicos del individuo, como gestos, expresiones faciales y patrones de voz.

Además, los avatares pueden personalizarse casi ilimitadamente. Pueden ajustar el estilo y la apariencia en función de la preferencia del individuo, lo que le permite seleccionar o cambiar atributos como el género, la apariencia física, la ropa y hasta el estado de ánimo. El uso de la IA para crear avatares permite a los usuarios crear identidades digitales en el metaverso de forma más rápida y sencilla. A través del avatar, uno puede mostrar su verdadera cara o probar alternativas. Generalmente, esta es una oportunidad para realizar un viaje de exploración de la identidad, que no siempre es realizable fuera del mundo digital.

Con los avatares de IA, las personas pueden demostrar lo que les gustaría ser, pero no tienen valentía o habilidades para exprimirse al máximo; por ejemplo, porque tienen una apariencia distinta, hablan de manera diferente o se comportan de otra manera. Además, la IA hace que estos avatares sean más reactivos y adaptables a los sentimientos. De hecho, la IA está siendo diseñada para que los avatares entiendan e interpreten las emociones de los otros y respondan en consecuencia. Esto, a su vez, hace que la comunicación sea más sincera. En efecto, muchas veces, las personas eligen la comunicación no verbal, como por ejemplo la comunicación corporal y las expresiones faciales, que en el metaverso enriquecen la comunicación humana y la hacen más real.

Por lo tanto, el metaverso, como se ha explicado por la IA, ha cambiado la forma en que interactuamos digitalmente y nos expresamos, ya que los avatares de IA reflejan nuestra identidad real al igual que nuestra identidad deseada. Por eso, a través de espacios personalizados y avatares más realistas, la IA hace del metaverso un espacio para la creatividad, sin límites y auténtico, creando un nuevo campo de interacción dentro del mundo virtual.

10. El *deepfake*, ¿puede ser delito?

La creación de un *deepfake* y su posterior difusión en la red o las redes sociales podría derivar en delito en ciertos casos. Como se ha mencionado en algunos análisis, la creación de un *deepfake* con la imagen de una persona por sí sola no puede considerarse un delito contra la intimidad o lesiones en la libertad de esos cuerpos (artículo 197 del Código Penal).

Si se trata de una imagen original, no generada artificialmente, podría constituir un delito acceder a imágenes, tanto publicadas como no publicadas, siempre que se violen las medidas de protección de datos diseñadas para salvaguardar la información personal. Sin embargo, es habitual que este tipo de imágenes sean extraídas de desarrollos o publicaciones accesibles al público, como perfiles en redes sociales y plataformas de acceso abierto. En otras palabras, provienen de fuentes públicas y no privadas. Esto plantea una nueva perspectiva en el análisis del delito, dado que las imágenes rara vez se obtienen directamente de fuentes privadas.

Dado que su creación se encuentra en la obra artificial, en un alto porcentaje de los casos se utilizan para humillar o desacreditar, generalmente ofendiendo la imagen de una persona. Por estos motivos, el delito más relevante que podría ser aplicable en este caso es el de injurias.

Como resultado, este contenido no solo perjudica a la reputación de la gente, sino que también causa daño emocional a las víctimas.

De acuerdo con el artículo 208 del Código Penal español, la injuria se refiere a cualquier «acción o expresión que lesionen la dignidad de otra persona, menoscabando su fama o menospreciando su propia estimación». Así, desde la perspectiva de la ley, esta disposición regula el derecho del honor, la reputación y la propia estimación de una persona. Desde esta perspectiva, los *deepfakes* pueden interpretarse como un tipo de lesión. Es apropiado decir que los *deepfakes*, cuando ciertas *deepfakes* dañan, son un tipo de lesión. El *deepfake* es la simulación necesaria con el *software* de una persona en la relación que no ha creado en realidad. Por lo tanto, tal representación es distorsionada y ofensiva. Solo los *deepfakes*, en su creación y, en menor medida, en su distribución, que modelan a una persona en relación con cualquier situación negativa o la convierten en un objeto de burla y humillación, pueden lesionar su dignidad y menospreciar su reputación y propia autopercepción.

Sin embargo, se debe destacar que no todas las lesiones son delitos. La tipificación penal actual está permitida para persecución de lesiones en casos de especial gravedad. Por lo tanto, para que un *deepfake* pueda ser calificado como delito de lesión, su propio contenido y la percepción pública de este contenido deben ser particularmente hirientes o dañinos para la reputación de la persona retratada. Nuevamente, el término «gravedad» se debe entender de manera contextual y tener en cuenta en las circunstancias del caso. En otras palabras, una persona debe ser capaz de comprender la intención del creador del *deepfake*, así como evaluar el potencial de daño a la reputación real de la víctima. Si la persona representada se muestra a la luz de tal manera que propaga la deshonra de su imagen, o la percepción del público está distorsionada para producir su daño, esto es daño grave, a saber, esto es lo que está tipificado en el delito de injuria. Si, por otro lado, la presentación no causara ningún daño real, o la presentación se vea como tonta, no habría nada que juzgar.

Para juzgar, la parte lesionada tendría que presentar una querella en el juzgado de instrucción. Se recomienda que la demanda, en caso de que sea civil, sea redactada por un abogado especializado en delitos contra el honor. Esta persona ayudará a preparar las pruebas necesarias, y así, basándose en los requisitos legales propuestos, maximizará las posibilidades de éxito en el tribunal. También informará al demandante sobre los recursos legales y compensatorios disponibles en la ley.

Al mismo tiempo, la mayoría de los estudios revisados argumentan sobre *deepfakes* predominantemente sexuales o pornográficos, indicando que «en estas circunstancias, el deepfake fue creado con el deseo de humillar e insultar por delante de la persona». Teniendo en cuenta la forma clara del estilo de su contenido, su naturaleza explícita y el daño real a la dignidad de la persona cuya imagen se utilizó en él, este estilo de *deepfake* debería ser reconocido como un delito en todos los aspectos. Se ha demostrado que el uso de la imagen de una persona para cualquier contenido sexual sin permiso está destinado a dañar el honor y la intimidad debido a la naturaleza, que suele ser negativa y, a menudo, falsa y degradante.

La divulgación de este contenido puede tener graves consecuencias destructivas. Dado que el hecho de exhibir esta forma de contenido puede conducir a un colapso del estado moral de una persona, la envidia, el dolor y la imposibilidad de trabajar o tener relaciones personales, este último es evidente. Por lo tanto, es urgente y necesario regular y abordar la cuestión de la ética y la legitimidad del *deepfake*.

Dado el refinamiento constante de las tecnologías que permiten crear este contenido, es evidente que el desafío de regular la distribución de este y proteger contra estos delitos se está volviendo cada vez más acuciante. Un marco legal mejorado permitiría ofrecer a las personas una mayor protección en relación con los *deepfakes* y ofrecer una defensa real contra los riesgos planteados por estas representaciones manipuladas.

Por todo ello, nos encontramos en un escenario que plantea serias dudas sobre la evolución que puede experimentar la regulación legal frente a fenómenos tan nuevos como el de los *deepfakes*.

Desde un punto de vista legal, ya se sabe que el Derecho es reactivo, es decir, que es una respuesta a los conflictos humanos para restablecer el equilibrio y garantizar los derechos violados. Pero, al mismo tiempo, ya que el Derecho es reactivo, la regulación siempre llega demasiado tarde para una determinada categoría de conflictos. Esto se debe a que el siglo

pasado vio surgir en los países más desarrollados una nueva ola de tecnologías que superaron la adaptabilidad de la ley y el reglamento a ellos.

Hoy en día, el marco legal en muchos países está desactualizado y, a menudo, no es del todo adecuado para resolver problemas modernos, como los delitos de *deepfakes* y otras manipulaciones digitales. Y no solo no protege a las víctimas, en su mayor parte, sino que también deja un hueco en forma de vacío normativo, en el que pueden encontrar refugio incluso aquellos que usan tecnologías de forma dañina. Esto se debe a que, además de la posible falta de una respuesta judicial clara sobre la viabilidad de manipular digitalmente las acciones, las personas tienden a recurrir a esta práctica cuando consideran que se les ha negado algo de manera injusta.

Hay varias razones para esta situación. En primer lugar, el factor fundamental es la falta de legislación sobre inteligencia artificial y uso tecnológico para hacer frente a los desafíos; en muchos casos, la legislación existente no permite que las solicitudes legales sobre delitos de falsificación y otras formas de abuso digital se hagan realidad. En segundo lugar, los tribunales en muchos casos carecen de un marco claro para tratar de abordar estos delitos; en tales condiciones, los jueces y fiscales quizás no cuenten con la formación adecuada para aplicar la ley de manera que se adapte a tal estado de cosas. Por lo tanto, los medios no estarán disponibles para que los tribunales impartan justicia a las víctimas y castiguen a los delincuentes de manera efectiva. En tercer lugar, falta de herramientas tecnológicas y recursos humanos para permitir a las organizaciones de seguridad y los tribunales investigar y castigar los delitos. Sin estos recursos, las persecuciones penales de delitos perpetrados no serán realizables; es extremadamente difícil rastrear el origen de una falsificación profunda, y con frecuencia imposible identificar al autor. En cuarto lugar, la falta de ajuste de la legislación a las necesidades actuales no es solo un problema, sino también una oportunidad crítica para modificar la forma en que la ley se aplica en el entorno digital. Establecer normas específicas, capacitar a las instituciones jurídicas y dotar a los órganos de justicia de las herramientas tecnológicas necesarias son pasos imprescindibles para que el Derecho pueda cumplir su función de protección y regulación en una sociedad cada vez más influenciada por el desarrollo tecnológico.

11. Conclusiones

El presente estudio ha revelado varias ideas clave sobre sus implicaciones éticas, sociales y legales. A continuación, se presentan algunas de las principales conclusiones derivadas de este análisis:

1. Explotación de la privacidad y consentimiento: La tecnología *deepfake* ha permitido la creación de contenidos explícitos sin el consentimiento de las personas, lo que plantea serias preocupaciones sobre la violación de la privacidad y otros derechos inherentes a toda persona. Las personas pueden ser representadas en situaciones sexuales o explícitas sin su aprobación, lo que vulnera su privacidad y expone a las víctimas a daños psicológicos y sociales graves.
2. Aumento de la pornografía no consensuada: Los avatares *deepfake* han dado lugar a un aumento de la pornografía no consensuada. Esto ocurre cuando imágenes o vídeos explícitos de una persona se crean sin su permiso y se difunden en plataformas en línea, lo que genera un daño considerable en términos de reputación, salud mental y bienestar de las víctimas.

3. Desafíos éticos y legales: La creación y distribución de pornografía *deepfake* genera un dilema ético complejo sobre la responsabilidad y el control de la tecnología. A pesar de que esta tecnología tiene aplicaciones legítimas en áreas como el entretenimiento y la educación, su uso malintencionado en la pornografía no consensuada plantea interrogantes sobre la necesidad de una regulación más estricta para proteger la dignidad y los derechos de las personas.
4. Impacto psicológico y social en las víctimas: Las víctimas de pornografía *deepfake* a menudo experimentan un daño psicológico significativo. Además, las repercusiones sociales de la difusión de estos contenidos afectan a la reputación de las personas, incluso cuando el contenido es claramente falso. Esto demuestra la necesidad de estrategias públicas de apoyo y recursos para ayudar a las personas afectadas.
5. Necesidad urgente de legislación y regulación: La regulación de la pornografía *deepfake* es una necesidad urgente para prevenir el abuso de la tecnología y proteger a las personas afectadas. Si bien algunos países han comenzado a implementar leyes contra la creación y distribución de contenido explícito no consensuado, aún existe una laguna significativa en la legislación global, lo que requiere esfuerzos coordinados para abordar de manera efectiva este problema.
6. Responsabilidad de las plataformas tecnológicas: Las plataformas en línea juegan un papel fundamental en la propagación de la pornografía *deepfake*, ya que muchos de estos contenidos se difunden rápidamente a través de redes sociales y sitios de alojamiento de vídeos. Las empresas tecnológicas deben asumir una mayor responsabilidad para implementar medidas más efectivas que detecten y eliminen este tipo de contenido, así como mejorar los protocolos de denuncia y apoyo a las víctimas.
7. Desafíos en la detección de *deepfakes*: La capacidad de detectar *deepfakes* sigue siendo un reto significativo debido a la sofisticación de la tecnología. Las técnicas de detección existentes no siempre son efectivas, lo que hace que la creación y distribución de este contenido malicioso sea aún más peligrosa. Esto resalta la necesidad de desarrollar mejores herramientas de verificación y de educación pública sobre cómo identificar este tipo de contenido.

La tecnología *deepfake* presenta tanto oportunidades como riesgos. Mientras que su uso legítimo puede enriquecer la industria del entretenimiento y otras áreas, su aplicación en la pornografía no consensuada plantea serias preocupaciones sobre la explotación, el daño psicológico y la violación de derechos. Para abordar estos desafíos, es esencial desarrollar marcos legales robustos, aumentar la responsabilidad de las plataformas tecnológicas y proporcionar apoyo adecuado a las víctimas. El estudio de los avatares *deepfake* y la pornografía resalta la necesidad de un enfoque integral para mitigar los riesgos asociados con el abuso de estas tecnologías y proteger los derechos fundamentales de las personas en la era digital.

En base a lo explicado en el presente artículo, la proliferación de tecnologías está planteando un desafío mayúsculo en materia de regulación legal, ya que tiene una evolución constante. La falta de un marco legal actualizado y adecuado deja desprotegidos tanto a las víctimas como a los autores, los cuales, hasta la fecha de este artículo, tienen un entorno donde poder cometer todas estas conductas delictivas. Solo se podrá garantizar la protección

de la dignidad o de los derechos de las personas en un entorno digital si, como sociedad, conseguimos adaptar nuestra legislación a la realidad social.

12. Bibliografía

Bach-Lombardo, J., & Winter, C. (2016, February 13). *Why Isis propaganda works*. The Atlantic. https://www.theatlantic.com/international/archive/2016/02/isis-propagandawar/462702/

Baele, S. J., Brace, L., & Coan, T. G. (2020, December 30). *Uncovering the far-right online ecosystem: An analytical framework and research agenda*. Taylor & Francis Online. https://www.tandfonline.com/doi/full/10.1080/1057610X.2020.1862895.

Brooks, T.; Princess, G.; Heatley, J.; Kim, S.; Parks, S.; Reardon, M., et al. (2019). *Increasing Threat of Deepfake Identities*. Department of Homeland Security. https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_ deepfake_identities_0.pdf.

Bunk, J.; Bappy, J. H.; Mohammed, T. M.; Nataraj, L.; Flenner, A.; Manjunath, B. S., et al., «Detection and localization of image forgeries using resampling features and deep learning», 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1881-1889, 2017, July.

Burkell, J., y Gosse, C. (2019). Nothing new here: Emphasizing the social and cultural context of deepfakes. *First Monday.* https://doi.org/10.5210/fm.v24i12.10287

Byman, D. L.; Gao, C.; Meserole, C., y Subrahmanian, V. S. (2023). *Deepfakes and international conflict*. Foreign Policy at Brookings. https://www. brookings.edu/wp-content/uploads/2023/01/FP_20230105_deepfakes_international_ conflict.pdf?mc_cid=3f7678e334&mc_eid=ee94395166.

Cabrera Schulmeyer, M. C. (2023). Metaverso, avatares, inteligencia artificial: ¿De qué estamos hablando realmente (¿o virtualmente?). *Revista Chilena de Anestesia, 52*.

Cauberghs, O. (2023, November 13). For the Lulz?: AI-generated subliminal hate is a new challenge in the fight against online harm. *Global Network on Extremism and Technology*. https://gnet-research.org/2023/11/13/for-the-lulz-ai-generated-subliminal-hate-is-a-newchallenge-in-the-fight-against-online-harm/

Codina, L. (2020a). Revisiones bibliográficas sistematizadas en Ciencias Humanas y Sociales. 1: Fundamentos. *Methodos Anuario de Métodos de Investigación en Comunicación Social, 1,* 50-60. https://doi.org/10.31009/methodos.2020.i01.05

Codina, L. (2020b). Revisiones sistematizadas en Ciencias Humanas y Sociales. 2: Búsqueda y Evaluación. *Methodos Anuario de Métodos de Investigación en Comunicación Social, 1,* 61-72. https://doi.org/10.31009/methodos.2020.i01.06

Codina, L. (2020c). Revisiones sistematizadas en Ciencias Humanas y Sociales. 3: Análisis y Síntesis de la información cualitativa. *Methodos Anuario de Métodos de Investigación en Comunicación Social, 1,* 73-87. https://doi.org/10.31009/methodos.2020.i01.07

Cook, J. (2022, July 27). Deepfake technology: Assessing security risk. *American University*.

Flores, M. (2022, May 31). The new world order: The historical origins of a dangerous modern conspiracy theory. *Middlebury Institute of International Studies at Monterey*. https://www.middlebury.edu/institute/academics/centers-initiatives/ctec/ctec-publications/new-worldorder-historical-origins-dangerous.

Flynn, A.; Powell, A.; Scott, A. J., y Cama, E. (2021). Deepfakes and digitally altered imagery abuse: A cross-country exploration of an emerging form of image-based sexual abuse. *The British Journal of Criminology, 62*(6), 1341-1358. https://doi.org/10.1093/bjc/azab111

Frankfurt, H. G. On Bullshit (Oxford: Oxford University Press, 2005). Google Scholar.

Franks, B.; Bangerter, A., y Bauer, M.W. Conspiracy theories as quasi-religious mentality: an integrated account from cognitive science, social representations theory, and frame theory. *Front Psychol.* 2013 Jul 16;4:424. doi: 10.3389/fpsyg.2013.00424. PMID: 23882235; PMCID: PMC3712257.

Fricker, M. Epistemic Injustice: power and the ethics of knowing (Oxford: Oxford University Press, 2007). CrossRefGoogle Scholar.

Gans, J. (2023, June 7). FBI warns of 'deepfakes' in sextortion schemes. *The Hill.* https://thehill.com/policy/cybersecurity/4037204-fbi-warns-of-deepfakes-in-sextortion-schemes/.

Gosse, C., y Burkell, J. (2020). Politics and porn: How news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication, 37*(5), 497-511. https://doi.org/10.1080/15295036.2020.1832697

Harper, C. A.; Fido, D., y Petronzi, D. (2019). Delineating non-consensual sexual image offending: Towards an empirical approach. *Sexual Abuse: A Journal of Research and Treatment, 31*(6), 706-725. https://doi.org/10.31234/osf.io/vpydn

Helmus, T. C. (2022, July 6). Artificial intelligence, deepfakes, and disinformation: A primer. *RAND Corporation*. https://www.rand.org/pubs/perspectives/PEA1043-1.html.

Jacobsen, B. N., y Simpson, J. (2023). The tensions of deepfakes. *Information, Communication & Society*, 1-15. https://doi.org/10.1080/1369118x.2023.2234980

Jarvis Cooper, L. (2022). Sexual Privacy and Persecution. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.4072440

Jolley, D., y Douglas, K. M. (2017). Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology,* 47(8), 459–469. https://doi.org/10.1111/jasp.12453

Karasavva, V., y Forth, A. (2021). Personality, Attitudinal, and Demographic Predictors of Non-consensual Dissemination of Intimate Images. *Journal of Interpersonal Violence, 37*(21-22), NP19265–NP19289. https://doi.org/10.1177/08862605211043586

Laffier, J., y Rehman, A. (2023). Deepfakes and Harm to Women. *Journal of Digital Life and Learning, 3*(1), 1-21. https://doi.org/10.51357/jdll.v3i1.218

Lucas, K. T. (2022). Deepfakes and Domestic Violence: Perpetrating Intimate Partner Abuse Using Video Technology. *Victims & Offenders, 17*(5), 647-659. https://doi.org/10.1080/15564886.2022.2036656

Lucas, K. T. (2023). Revenge Porn and Deepfake Technology: A Domestic Violence Perspective. *Journal of Interpersonal Violence, 38*(7), 1-23.

Mania, K. (2022). Legal Protection of Revenge and Deepfake Porn Victims in the European Union: Findings From a Comparative Legal Study. *Trauma, Violence & Abuse, 25*(1), 117-129. https://doi.org/10.1177/15248380221143772

Martínez Sánchez, M. (2023). El discurso sobre el revenge porn en la prensa: estudio de caso de Rosalía y sus fotografías manipuladas. *Journal of Feminist, Gender and Women Studies, 15*, 94-115. https://doi.org/10.15366/jfgws2023.15.005

Okolie, C. (2023). Artificial intelligence-altered videos (deepfakes), image-based sexual abuse, and data privacy concerns. *Journal of International Women's Studies, 25*(2), 11.

O'Sullivan, D.; Devine, C., y Gordon, A. (2023, November 15). How antisemitic hate groups are using artificial intelligence in the wake of Hamas attacks. *CNN*. https://www.cnn.com/2023/11/14/us/hamas-israel-artificial-intelligence-hate-groups-invs/index.html.

Paris, B., y Donovan, J. (2019). Deepfakes and cheap fakes. *Data & Society.* https://datasociety.net/library/deepfakes-and-cheap-fakes/ (acceso 5 de diciembre de 2023)

Pearson, J., y Zinets, N. (2022, March 17). Deepfake footage purports to show Ukrainian president capitulating. *Reuters*. https://www.reuters.com/world/europe/deepfakefootage-purports-show-ukrainian-president-capitulating-2022-03-16/.

Quirk, C. (2023, June 19). The high stakes of deepfakes: The growing necessity of federal legislation to regulate this rapidly evolving technology. *Princeton University*. https://legaljournal.princeton.edu/the-high-stakes-of-deepfakes-the-growing-necessity-of-federallegislation-to-regulate-this-rapidly-evolving-technology/.

Rousay, V. (2023). Sexual Deepfakes and Image-Based Sexual Abuse: Victim-Survivor Experiences and Embodied Harms (Doctoral dissertation, Harvard University).

Roy, R.; Dixit, A. K.; Saxena, S., y Memoria, M. (2023). Meta-analysis of artificial intelligence solution for prevention of violence against women and girls. In *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*. https://doi.org/10.1109/icicat57735.2023.10263765

Sayler, K. M., y Harris, L. A. (2023, April 17). Deep fakes and national security. *Congressional Research Service*. https://www.documentcloud.org/documents/23798946- deep-fakes-and-national-security-april-17-2023

Schmidt, B. (2023, October 3). Proposed Wisconsin bill to address artificially made «deep fake» pornography. *WBay*. https://www.wbay.com/2023/10/02/proposed-wisconsin-bill-addressartificially-made-deep-fake-pornography/.

Sharpe, M., y Mead, D. (2021). Problematic Pornography Use: Legal and Health Policy Considerations. *Current Addiction Reports, 8*(4), 556-567. https://doi.org/10.1007/s40429-021-00390-8

Siegel, D., y Chandra, B. (2023, August 29). «Deepfake Doomsday»: The role of artificial intelligence in amplifying apocalyptic Islamist propaganda. *Global Network on Extremism & Technology*. https://gnet-research.org/2023/08/29/deepfake-doomsday-the-role-ofartificial-intelligence-in-amplifying-apocalyptic-islamist-propaganda/

Simón Soler, E. (2023). Retos jurídicos derivados de la inteligencia artificial generativa. *InDret.* https://doi.org/10.31009/indret.2023.i2.11

United States Senate Committee on the Judiciary. (2023, June 13). *Artificial intelligence and human rights*. https://www.judiciary.senate.gov/committee-activity/hearings/artificial-intelligenceand-human-rights.

Van der Nagel, E. (2020). Verifying images: deepfakes, control, and consent. *Porn Studies, 7*(4), 424-429. https://doi.org/10.1080/23268743.2020.1741434

Walker, K., y Sleath, E. (2017). A systematic review of the current knowledge regarding revenge pornography and non-consensual sharing of sexually explicit media. *Aggression and Violent Behavior, 36,* 9-24. https://doi.org/10.1016/j.avb.2017.06.010

Warren, J. (2023, November 11). Fake audio of Sadiq Khan is not a crime, says Met. *BBC News*. https://www.bbc.com/news/uk-england-london-67389609.

Weiner, D. I., y Norden, L. (2023, December 12). Regulating AI deepfakes and synthetic media in the political arena. *Brennan Center for Justice*. https://www.brennancenter. org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena

Williams, K. (2023, October 24). Tightening restrictions on deepfake porn: What US lawmakers could learn from the UK. *Tech Policy Press*. https://techpolicy.press/tighteningrestrictions-on-deepfake-porn-what-us-lawmakers-could-learn-from-the-uk/

Yang, X.; Li, Y., y Lyu, S. «Exposing Deep Fakes Using Inconsistent Head Poses», ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 8261-8265, doi: 10.1109/ICASSP.2019.8683164.