Universitat Autònoma de Barcelona

Tipificación de las *deepfakes* sexuales en el Código Penal Español

Helena Araúz Mateu

DOBLE GRADO EN DERECHO Y RELACIONES LABORALES

Tutor: Joan Baucells Llados

Facultad de Derecho

Fecha: 6 de mayo de 2024

Resumen

El presente trabajo tiene por objeto hacer un análisis legal dentro del limitado marco normativo español frente a los riesgos que podría ocasionar una nueva tecnología disruptiva llamada *deepfake* o falsificación profunda, cuestionando si se podrían adaptar las categorías legislativas actuales a la nueva era de delitos cometidos a través de la inteligencia artificial. Para ello, se parte de una comprensión amplia de la naturaleza y dimensión de los cambios experimentados en la era digital, proponiendo una definición de *deepfake*.

En este contexto, esta investigación propone analizar la complejidad de incorporar los delitos perpetrados mediante *deepfakes* en un tipo penal específico en el sistema legal español. Para ello, la revisión de artículos doctrinales y jurídicos proporcionará un fundamento teórico sólido para comprender los desafíos que enfrenta el ordenamiento jurídico en la adaptación a la rápida evolución de las tecnologías de manipulación digital.

Una vez realizado dicho estudio, se deduce que, de acuerdo con el Código Penal actual, este tipo de delitos podrían subsumirse en el artículo 197.7 de revelación de secretos, el 173.1 de delitos contra la integridad moral y el 208 del delito de injurias. A lo largo del marco teórico, se realizará un análisis de cada uno de los tipos y la posibilidad de sumisión de los delitos cometidos por *deepfakes* sexuales en estos. En última instancia, también se realizará una exposición del derecho comparado, para examinar como otros ordenamientos y legislaciones están actuando frente a dicha tecnología.

Palabras clave: *deepfake*, inteligencia artificial, fake news, deep learning, GAN (red de acceso genérico).

Abstract

The aim of this work is to carry out a legal analysis within the Spanish regulatory framework against the risks that could be caused by a new disruptive technology called *deepfake*, as well as any other type of AI. Questioning whether current legislative categories could be adapted to the new era of crimes committed through artificial intelligence. To do this, we start from a broad understanding of the nature and dimension of the changes experienced in the digital era, proposing a definition of *deepfake*.

In this context, this work aims to analyze the complexity of incorporating crimes perpetrated through *deepfakes* into a specific criminal type in the Spanish legal system. To this end, the review of doctrinal and legal articles will provide a solid theoretical foundation to understand the challenges faced by the legal system in adapting to the rapid evolution of digital manipulation technologies.

Once this study was carried out, it is deduced that according to the current Penal Code, this type of crimes could be subsumed in article 197.7 of disclosure of secrets, the 173.1 crimes against moral integrity and 208 the crime of libel. Throughout the theoretical framework, an analysis will be carried out of each of the types and the possibility of submission of crimes committed by sexual *deepfakes* in them. Ultimately, a presentation of comparative law will be made to examine how other systems and legislation are acting in relation to said technology.

Palabras clave: *deepfake*, artificial inteligence, fake news, deep learning, GAN (Generative Adversarial Network).

Índice

1. Introducción

Tal y como decía Hart (1994), el derecho, no puede ser estático, requiere de una continua transformación. El derecho por lo tanto, es una disciplina que busca adaptarse a la realidad; esto no quiere decir que las normas jurídicas deban ser interpretadas en relación a la realidad del tiempo en el que han de ser aplicadas, como expone el artículo 3.1 del Código Civil. Es decir, muchos preceptos deben ser pensados y analizados acorde al tiempo en que se deben aplicar, diferentes al momento en que se plantearon, pero con el objetivo de dar respuesta a problemas o situaciones actuales.

En la era digital contemporánea, la proliferación de tecnologías de manipulación de medios ha suscitado preocupaciones significativas en diversos ámbitos, incluido el ámbito legal y social. Entre estas tecnologías, los *deepfakes* han emergido como una herramienta poderosa y controvertida, capaz de generar contenido audiovisual falso con una calidad cada vez más convincente. Si bien los *deepfakes* tienen aplicaciones potenciales en campos como el entretenimiento y la publicidad, también plantean serios desafíos éticos y jurídicos.

Es por ello, que, ante un fenómeno nuevo, complejo y en expansión como son las *deepfakes*, se requieren ajustes importantes o, más bien, una reconceptualización de la legislación actual, ya que en el Código Penal vigente no hay ningún tipo de tipificación expresa a esta nueva tendencia que ha venido para quedarse.

En los últimos años, el desarrollo de estas tecnologías ha sido exorbitante, hasta el punto de que esta se ha convertido en una herramienta de uso cotidiano para un gran número de personas, como sería el caso de Chatgtp: un sistema de chat con inteligencia artificial entrenada para mantener conversaciones, lanzado en 2022. Así como, Copilot que es un sistema de inteligencia artificial creado por Microsoft basado en un sistema de chat. Pero junto a estas, se han desarrollado otras inteligencias artificiales denominadas *deepfakes*, una amalgama de las palabras "deep learning" y "fake". Estas creaciones algorítmicas, impulsadas por el poder de la inteligencia artificial, han demostrado ser herramientas de doble filo al poner en jaque la integridad y seguridad de la sociedad, ya que permiten la simulación, tanto en video como en fotografías, la imagen, voz o gestos de las personas, haciendo ver al espectador que alguien está haciendo o diciendo algo que, en realidad nunca hizo o dijo.

Estas creaciones digitales, aunque inicialmente afectaron a celebridades, su alcance se ha extendido a personas comunes, especialmente mujeres. Entre 2022 y 2023, la cantidad de pornografía *deepfake* creada aumentó de manera alarmante, pasando de 3725 videos en 2022 a 21019 en 2023 (Magisnet, 2023). Estas aplicaciones se basan en sistemas de inteligencia artificial generativos que transforman imágenes inocentes en desnudos simulados. La

facilidad de acceso y la rapidez con la que se pueden crear estos contenidos plantean serias preocupaciones.

La creación de *deepfakes* sexuales no está contemplada como un delito específico en el Código Penal español. Sin embargo, desde la perspectiva de la violencia de género, estas imágenes constituyen una forma de agresión (Magisnet, 2023), ya que se desinforma, se manipula y se destruye la reputación de las víctimas, en consecuencia, las víctimas acaban sufriendo problemas psicológicos y sociales, incluso si las imágenes no son reales.

El presente trabajo se propone explorar exhaustivamente las posibles vías de sumisión de las *deepfakes* sexuales dentro del marco legal español. A medida que las tecnologías de manipulación de medios continúan evolucionando y refinándose, es crucial entender cómo se pueden abordar los abusos y las violaciones de derechos que surgen de su mal uso.

El trabajo se estructurará en torno a tres áreas principales: en primer lugar, se examinarán los conceptos legales existentes en España que podrían aplicarse a las *deepfakes* de contenido sexual, como las injurias, la integridad moral y el descubrimiento y revelación de secretos. En segundo lugar, se analizarán los desafíos específicos que surgen al intentar encajar las *deepfakes* dentro de estos marcos legales. Finalmente, se explorarán posibles enfoques de política legislativa y reguladora que podrían adoptarse para abordar más eficazmente este problema en el futuro. Así como, un estudio de la legislación de otros países como Estados Unidos o Reino Unido que nos establezcan líneas a seguir a la hora de tipificar dichos delitos.

En un momento en que la tecnología avanza a un ritmo vertiginoso, es imperativo que la legislación y la sociedad se adapten para hacer frente a los nuevos desafíos que surgen en el ciberespacio, incluida la amenaza representada por las *deepfakes* sexuales.

2. Marco Teórico

2.1 Definición de los falsos audiovisuales (*deepfakes*): abordaje multidimensional

Para definir el concepto del fake audiovisual hay que recurrir a su morfología lingüística. La etimología de "*falso*" (fake, en inglés) proviene del latín falsus y significa "*engañar o engañarse a sí mismo*"(Cerdán & Padilla, 2019). La palabra "*audiovisual*" está formada por raíces latinas y significa "*relativo a la vista y el oído*". Sus componentes léxicos son audio y videre (escuchar y ver). Por lo tanto, cualquier obra catalogada como fake audiovisual sería aquella que pretende engañar a través de estímulos visuales y auditivos. Básicamente, un falso documental es un texto ficticio que utiliza las técnicas y los códigos del documental para generar en el espectador la sensación de que la historia que se está narrando es cierta (Roscoe & Hight, 2001).

Según la definición teórica acuñada por Hao Li y otros en su artículo seminal "*Deep Learning for Real-Time Digital Makeup*"(Alzahrani, Al-Bander, & Al-Nuaimy, 2021), los *deepfakes* son productos de técnicas de aprendizaje profundo que permiten la transferencia de atributos faciales y expresiones entre individuos en tiempo real. Esta definición inicial se centra en la manipulación facial, pero es solo el punto de partida para comprender la amplitud de los *deepfakes* en la actualidad.

La profundización en la naturaleza multifacética de los *deepfakes* nos lleva a considerar la intersección de la tecnología y la ética. Danielle Citron, en su obra "*Hate Crimes in Cyberspace"*(Citron, 2014), destaca que los *deepfakes* no son simplemente herramientas de entretenimiento, sino que también se utilizan con propósitos maliciosos, incluidos los delitos contra la libertad sexual. Esta perspectiva resalta la urgencia de abordar las implicaciones éticas y legales que rodean a los *deepfakes*, no solo desde una perspectiva técnica, sino también desde una mirada que reconozca las consecuencias sociales y psicológicas de su mal uso.

En el ámbito normativo, la definición y comprensión de los *deepfakes* se ve moldeada por el marco legal en constante evolución. La Unión Europea, en su Reglamento General de Protección de Datos (Reglamento (UE) 2016/679 , de 27 de abril de 2016), aborda la cuestión de la privacidad y la protección de datos personales, estableciendo principios que buscan salvaguardar la integridad de la información personal frente a posibles manipulaciones como las perpetradas mediante *deepfakes.* La inclusión de tales disposiciones normativas refleja el

reconocimiento de la importancia de preservar la privacidad en un mundo donde la tecnología desafía constantemente los límites de la autenticidad.

La definición de *deepfakes*, por ende, no puede ser encapsulada únicamente en términos técnicos, sino que debe abordarse desde una perspectiva holística que abarque los aspectos éticos, legales y tecnológicos. Como señala el informe del Instituto Brookings sobre *deepfakes* y democracia (Villasenor, 2019), la comprensión de este fenómeno requiere una sinergia entre la comunidad científica, los legisladores y los defensores de los derechos civiles para desarrollar enfoques integrales que mitiguen sus impactos negativos. Abordar esta problemática de manera integral implica no solo comprender la ingeniería detrás de los *deepfakes,* sino también analizar críticamente su impacto en la sociedad y las medidas regulatorias necesarias para preservar los derechos individuales y la integridad de la información en la era digital.

2.2 Aplicaciones y usos legítimos de las *deepfakes*

A pesar de que las implementaciones negativas son bastantes, también se han de considerar las positivas. Entre los usos beneficiosos de las *deepfakes*, se están implementando estas en el ámbito médico, por ejemplo, existe un estudio para un software que pueda de manera artificial regenerar la voz de las personas que no son capaces de hablar por culpa de alguna enfermedad como la esclerosis lateral amiotrófica (Citron & Chesney, 2019). Técnicas similares también podrían tener potencial para ayudar a gente en el proceso de luto, permitiéndole enzarzarse en una conversación de manera virtual con los seres queridos que ya no están (Meskys, Liaudanskas, Kalpokiene & Jurcys, 2020).

En 2019 se advierte ya una ligera expansión de géneros discursivos del *deepfake* en áreas relacionadas con arte, como por ejemplo en las obras de Klingemann (Sotheby´s, 2019) quien construye una galería en Twitter con el hashtag #BigGAN. Además, se empiezan a crear retratos vivientes o modelos de cabezas parlantes (talking heads) mediante tecnología de red neuronal convolucional impulsada por Samsung (Zakharov, Shysheya, Burkov & Lempitsky, 2019).

También se pueden utilizar las *deepfakes* para incrementar el atractivo de los museos, de las lecciones de historia, trayendo de manera artificial a la vida figuras históricas con fines educacionales (Citron & Chesney, 2019). Un ejemplo de ello es en el museo de Salvador Dalí en Sant Petersburgo, Florida, donde tienen una exhibición en que se muestra al artista. Los visitantes del museo pueden interactuar con él, así como ver a Dalí reencarnando sus frases célebres (Citron & Chesney, 2019). Otro ejemplo es que los investigadores de Samsung en Moscú crearon una *deepfake* de la Mona Lisa, es decir, usando el famoso cuadro,

utilizaban las facciones de otra persona para convergerlas en la cara de la Mona Lisa y crear videos de ella hablando (2019).

Los géneros discursivos de entretenimiento atraviesan y fusionan prácticas de la sátira y la parodia en memes deepfake virales como los de Obama sings Baka Mitai (Mechanical, 2020), Avengers (Ment, 2020). y una extensa colección de otros personajes célebres asociados a esta canción (Hao, 2020); o bien, en usos en videojuegos donde el usuario puede insertarse como avatar en tiempo real, como propone la plataforma Unreal Engine (Culture Shock, 2020). También, emergen otros géneros de entretenimiento en *deepfake*, como el espectáculo de cabaret creado por The Zizi Show (2020); y las transmisiones deportivas realizadas con sistemas de inteligencia artificial como Pixellot.tv (2021).

Para 2021 las técnicas de doblaje y traducción han evolucionado en el campo cinematográfico con el *deepfake*. Un ejemplo es Flawless, un programa realizado con la técnica de neural style-Preserving visual dubbing, que permite hacer un doblaje en diversos idiomas sin perder los rasgos originales de los actores (Vincent, 2021).

Entre 2020-2021 se encuentra como tendencia la mayor expansión de géneros discursivos con usos benéficos. Los usos del *deepfake* en géneros como el activismo político, campañas sociales y denuncia se han amplificado en este período. Por ejemplo, David Beckham apareció en un anuncio sobre la importancia de la salud, informando a las personas sobre la importancia de prevenir la malaria. La campaña denominada “Speak Up” se trataba de un video donde se le podía ver hablando en diferentes idiomas de manera fluida. Para hacerlo, se utilizaron *deepfakes* que modifican la forma de la boca, haciendo posible que David Beckham pudiese hablar diferentes idiomas (Meskys, Liaudanskas, Kalpokiene, & Jurcys, 2020).

Finalmente, recalcar que las *deepfakes* también se pueden utilizar para mejorar la privacidad de las personas. Ya que permiten a estas expresarse de manera anónima, así como la de preservar la identidad individual online. Un ejemplo de esto es que, gracias a este tipo de tecnologías, las personas que forman parte del colectivo LGTB que sufrían persecuciones en Rusia fueron capaces de testificar sus experiencias en el documental “Bienvenidos a Chechenia” de manera que no se les pudiese reconocer (RD, 2020).

2.3 Aplicaciones y usos ilegítimos de las *deepfakes*

Tal y como se ha discutido, la tecnología de las *deepfakes* pueden ser usadas en diferentes áreas. Y, estás, aunque tengan numerosos beneficios, como cualquier tecnología, vienen con una serie de riesgos.

Los avances en la tecnología han permitido a las *deepfakes* tener un gran impacto en el mundo. En los últimos años, con el fenómeno de las "*Fake News*", que se basa en la manipulación de un reportaje o de las noticias mediante la imitación del estilo y presentación, usando las *deepfakes*, para dar a entender que algo es cierto. Un ejemplo de ello, son los videos fakes realizados con los presidentes Trump y Obama emitiendo declaraciones falsas relacionadas con la cumbre de Davos (2020). También destacan los casos de Boris Johnson (Advocacy, 2019) y Jeremy Corbyn (Advocacy, 2019) donde emiten opiniones contrarias a sus idearios, o la Reina Isabel II grabando un video para TikTok en un mensaje de fin de año (Central Comedy, 2021).

En 2018, el presidente de Gabón, Ali Bongo Ondimba, se encontraba fuera del país recibiendo un tratamiento médico (Gabón, 2018). En este contexto, la gente empezó a preguntar porque no había hecho ninguna aparición pública, salvo algunas fotografías suyas publicadas por el gobierno junto con un vídeo mudo. En medio de las especulaciones, los asesores del presidente prometieron que pronunciaría su habitual discurso de Año Nuevo. Pero cuando el gobierno de Gabón publicó el vídeo, algunos conciudadanos, pensaron que había pocas dudas sobre la salud de su presidente. Mientras que otros, en concreto, el ejército de Gabón intentó un golpe de estado que finalmente fracasó citando la rareza del video como prueba de que algo andaba mal con el presidente (Breland, 2019). Por lo tanto, tal y como adelantaban Chesney y Citron (2019) estas *deepfakes* pueden ser utilizadas en el contexto de conflictos armados.

Cabe destacar que no solo pueden provocar problemas a nivel internacional, sino también a nivel doméstico. Diakopoulos y Johnson (2020) realizan varias formas en las que se pueden utilizar las *deepfakes* para alterar el procedimiento electoral, ya sea mediante la creación de videos que muestran a un candidato haciendo comentarios sexistas o racistas, así como audios que sugieren que un candidato tenía conocimiento a priori sobre alguna pregunta del debate público, etc. Fletcher (2018) muestra formas más sutiles en las que se pueden utilizar las *deepfakes* para manipular, por ejemplo, los resultados electorales, como a través del "*crowd-turfing*", es decir, hacer ver que el candidato tiene más o menos apoyo popular del que realmente tiene.

En una sociedad donde la información no solo se consume, sino que también se reproduce a un ritmo exorbitante en diferentes foros online, las *deepfakes,* pueden tener un efecto destructivo, ya que los videos o las imágenes alteradas suelen permanecer durante largos periodos de tiempo en internet.

En el reciente estudio elaborado por Europol (2023) se han identificado algunas utilizaciones de sistemas de IA generativa que pueden fomentar, facilitar o mejorar la comisión de determinados tipos delictivos. En primer lugar, definiremos la ciberdelincuencia de género

como aquellos delitos cometidos a través de Internet por razón de género prevaliéndose el agresor del alcance y la especial lesividad de los medios tecnológicos, tanto en el ámbito público como en el ámbito privado, con independencia de la relación preexistente con la víctima (Gónzalez, 2017).

En segundo lugar, destacar que estas funciones podrían mejorar y favorecer las prácticas de algunos ciberdelitos e incluso utilizar la posible generación de vídeo y audio para engañar o embaucar a menores o mujeres con el fin último de captar víctimas de otros eventuales delitos o bien directamente con el objetivo de perpetrar un delito sexual en el medio físico offline (Gónzalez, 2023).

2.4 Funcionamiento y evolución de las *deepfakes*

Las *deepfakes* se pueden definir como una combinación de aprendizaje profundo y falso. Los *deepfakes* son vídeos hiperrealistas manipulados digitalmente para representar a personas que dicen y hacen cosas que en realidad nunca dijeron ni sucedieron. Los *deepfakes* se basan en redes neuronales que analizan grandes conjuntos de datos para aprender a imitar las expresiones faciales, los gestos y la voz de una persona. El proceso implica introducir imágenes de dos personas en un algoritmo de aprendizaje profundo para intercambiar caras (Rösser, Cozzolino, Verdoliva & Riess, 2018).

Los recientes avances en las nuevas tecnologías de edición de audio, videos e imágenes, así como las herramientas que permiten dicha edición, tales como las GAN (Generative Adversarial Networks) (Lee, 2018) y, las redes sociales permiten la creación y rápida distribución de videos, imágenes y fotos manipuladas. Las GAN son una clase de algoritmos de IA que se utilizan en el aprendizaje no supervisado, implementados por un sistema de dos redes neuronales que compiten mutuamente. Esta técnica, presentada por investigadores de la Universidad de Montreal (Goodfellow, Pouget-Abadie, Mirza y otros, 2014), puede generar fotografías que parecen auténticas. Las GAN se basan en dos redes neuronales artificiales que trabajan juntas para crear medios de aspecto real. Estas dos redes llamadas el generador y el discriminador contienen el mismo conjunto de datos de imágenes, vídeos o sonidos. Es por ello, que mientras el primero intenta crear nuevas muestras que sean lo suficientemente buenas para engañar a la segunda red, esta intenta determinar si los nuevos medios que ve son reales.

En el caso de los *deepfake*, el generador de las GAN crea nuevas imágenes basándose en una base de datos de fotografías o vídeos. Una vez el algoritmo generador crea el nuevo contenido, el algoritmo discriminador realiza un cribado sobre el resultado generado, para delimitar si la imagen o vídeo reúne las características para ser un *deepfake* realista.

El uso de las GAN permite el continuo desarrollo, ya que se pueden pasar por el mismo proceso varias veces, permitiendo al discriminador rechazar cualquier material que pueda exponer que dicho contenido se trata de una deepfake. Este proceso, permite aprender de los errores y enmendarlos, para crear nuevas *deepfakes* que sean incluso más difíciles de detectar. Ya que contra más sofisticado es el proceso de producción y los resultados adquiridos usando las técnicas del "*deep learning*", permite distinguir las *deepfakes*, de otras técnicas de manipulación de datos, audio, etc llamadas "*cheap fakes*" (Paris, 2021).

La técnica de las *deepfakes* se creó mediante herramientas y funciones que son de acceso público generadas por grandes compañías como NVida y Google (Chawla, 2019). Esto significa que mientras el conocimiento técnico y el entendimiento de los parámetros de los programas son necesarios para desarrollar tal técnica, el software está disponible para el público general. Por ende, existe una gran amenaza por parte de las *deepfakes*, ya que, tal y como establece la DARPA (Defense Advanced Research Project Agency) una agencia del departamento de defensa de Estados Unidos, incluso una persona sin talento puede manipular cualquier medio visual (Siekierski, 2019).

Es por ello, que sería posible para cualquier persona con habilidades rudimentarias con el ordenador crear *deepfakes* de contenido sexual en las que se retrata a un individuo envuelto en un acto sexual que nunca tuvo lugar. A pesar de que, el origen de las *deepfakes* fue para crear imágenes, videos o audios falsos sobre famosos, cada vez más, estos se están utilizando para crear contenidos sobre personas cotidianas como un amigo, compañero o conocido.

De acuerdo con el informe elaborado por Deeptrace Labs (2019), publicado en el año 2019, se estableció que la mayoría de los vídeos *deepfake* eran vídeos de contenido sexual, un 96%, frente al 4% que no lo eran. Además, destacaron que en este tipo de vídeos las protagonistas eran mujeres, frente a los vídeos que no tienen contenido sexual donde los protagonistas eran hombres (Deeptrace Labs, 2019). En estos, tal y como se ha mencionado anteriormente, se utilizan técnicas de intercambio de caras para producir el contenido gráfico. Pero lo más preocupante, es que cada vez más son los casos que surgen de personas anónimas que son utilizadas para dichos fines.

En este mismo estudio, se expone que para hacer un video de contenido sexual en la *deepfake* de 60 segundos, hacen falta menos de 25 minutos sin coste alguno, aunque también existen de pago, usando solo una imagen del rostro de alguien, según datos del informe State of *deepfakes* de Home Security Heroes (2023). De acuerdo con este estudio, entre 2022 y 2023 la cantidad de *deepfakes* sexuales creada aumentó un 464%, pasando de 3725 vídeos en 2022 a 21019 en 2023.

Un ejemplo más reciente, lo encontramos en 2017 en el que un usuario anónimo con el seudónimo "*deepfakes*" publicó un video de contenido sexual en Reddit vanagloriándose que pertenecían a actrices famosas como Scarlett Johanson, Aubrey Plaza, Maisie Williams, etc. A pesar de que, estos videos se eliminaron rápidamente, esta técnica de reemplazo facial basada en la *Deep learning*, ganó la atención de los medios y se propagó por los foros. Del mismo modo, compartió un código abierto para *deep learning* en bibliotecas populares y así comenzaron a proliferar *deepnudes* de manera ilegal en diversos sitios (Adriani, 2019).

Un ejemplo más reciente de este suceso de las *deepfakes* sexuales es que en 2021 en canales de Telegram "desnudan" a una mujer a partir de una foto, siguiendo los pasos de la aplicación *DeepNude,* cerrada en 2019. Una práctica que algunos autores han denominado Automate Abuse Image (Abuso automatizado de imagen) y que afectó a más de 100 mil mujeres en 2020 (Adjer, Patrini & Cavalli, 2019). En 2018, por ejemplo, se observa un caso de porno-extorsión hacia la periodista Rana Ayyub (2018), acosada mediante un video sexual *deepfake,* con el fin de hacerla callar por una denuncia política realizada por ella, siendo víctima de una campaña de desinformación y desprestigio en redes digitales (Ayyub, 2018). En ese mismo año, Jordan Peel publica ¡You Won't Believe What Obama Says In This Video! en BuzzFeed (Silverman, 2018), en dicho video, tal y como se ha comentado anteriormente, se advierte de los riesgos de desinformación que pueden generar los *deepfakes*.

3. Abordaje legal de las *Deepfakes* de contenido sexual

3.1 Normativa Europea

Los esfuerzos legislativos por regular la IA en la UE comienzan en abril de 2018 con la comunicación: "Artificial Intelligence for Europe" a manos de la Comisión Europea (COM/2018/237 de 25 abril de 2018). El objetivo principal de la misma es adaptar la legislación europea a una realidad cambiante y en auge del desarrollo de las nuevas tecnologías. Ese mismo año, en diciembre, se emite una nueva comunicación conocida como el "Plan coordinado sobre la inteligencia artificial" (COM/2018/795, de 7 de diciembre de 2018), cuya finalidad es trabajar juntamente con los estados miembros para fomentar la cooperación en la UE y establecer una dirección común a seguir.

En abril de 2019 la Comisión realiza una nueva comunicación: "Generando confianza en una Inteligencia Artificial centrada en el ser humano" (COM/2019/16, de 8 de abril de 2019), cuyo propósito es el de involucrar a las partes interesadas en el desarrollo de la IA, centrándose en la ética y valores europeos.

El proceso legislativo en esta materia encuentra su culmen a nivel europeo con el acuerdo provisional de Reglamento sobre Inteligencia Artificial (COM/2021/206, 21 de abril de 2021) que es el instrumento jurídico mediante el que se pretende regular la aplicación de la IA a nivel europeo, esta ley tiene su origen en la Propuesta de la Comisión Europea del 2021. La Ley tiene dos objetivos claros:

1. Garantizar que los sistemas de Inteligencia Artificial utilizados en la Unión Europea e introducidos en el mercado europeo sean seguros y respeten los derechos de los ciudadanos.

2. Estimular la inversión y la innovación en el ámbito de la IA en Europa. El acuerdo establece que el Reglamento de Inteligencia Artificial debe aplicarse dos años después de su entrada en vigor.

Tal y como se establece en la exposición de motivos segunda del Reglamento, la base jurídica de la propuesta es, en primer lugar, el artículo 114 del Tratado de Funcionamiento de la Unión Europea (TFUE), que trata de la adopción de medidas para garantizar el establecimiento y el funcionamiento del mercado interior.

Para conseguir este objetivo, el Reglamento nos hace una clasificación de los diferentes sistemas de IA en cuatro niveles de riesgos, y en función de estos, se implementará un mayor o menor número de obligaciones. La división que nos presenta está basada en sistemas de IA que conllevan un riesgo inaceptable, un riesgo alto, un riesgo bajo o uno mínimo, estos últimos pueden elegir si se adhieren a sistemas voluntarios de cumplimiento. La lista de prácticas prohibidas abarca todos los sistemas de IA cuyo uso se considera inaceptable por ser contrario a los valores de la Unión, estos incluyen:

- manipulación cognitiva del comportamiento de personas o grupos vulnerables específicos

- puntuación social: clasificación de personas en función de su comportamiento, estatus socioeconómico o características personales

- sistemas de identificación biométrica en tiempo real y a distancia

El catalogado como de riesgo alto es aquel que produce un riesgo a los derechos y libertades de los individuos, esto implica que se verá sometido al cumplimiento de una serie de requisitos obligatorios, así como medidas de evaluación. La Propuesta de Reglamento califica como sistemas de alto riesgo, según lo establecido en la exposición de motivos 5.2.3, aquellos que acarrean un alto riesgo para la salud y la seguridad o los derechos fundamentales de las personas físicas (COM/2021/206, 21 de abril de 2021). Es por ello que están permitidos en el mercado europeo siempre que cumplan determinados requisitos obligatorios y sean sometidos a una evaluación de la conformidad *ex ante*. Estos se dividirán en dos categorías:

- Los sistemas de IA que se utilicen en productos sujetos a la legislación de la UE sobre seguridad de los productos, que incluye juguetes, aviación, automóviles, dispositivos médicos y ascensores.

- Los sistemas de IA pertenecientes a ocho ámbitos específicos que deberán registrarse en una base de datos de la UE:

    - Identificación biométrica y categorización de personas físicas

    - Gestión y explotación de infraestructuras críticas

    - Educación y formación profesional

    - Empleo, gestión de trabajadores y acceso al autoempleo

- Acceso y disfrute de servicios privados esenciales y servicios y prestaciones públicas

- Aplicación de la ley

- Gestión de la migración, el asilo y el control de fronteras

- Asistencia en la interpretación jurídica y aplicación de la ley.

Deben ir acompañados de un sistema de gestión de riesgos (art.9), una adecuada práctica en materia de gestión y gobernanza de datos (art.10), exigencia de documentación técnica (art.11), registro de la actividad de los sistemas de IA (art.12), transparencia y comunicación de información a los usuarios (art.13), vigilancia humana (art.14) y precisión, solidez y ciberseguridad (art.15).

En resumen, lo que se busca es evaluar el impacto que tiene cualquier nuevo sistema de IA en los derechos fundamentales antes de ser introducido en el mercado.

En octubre del 2020 el Parlamento Europeo aprobó la Resolución con recomendaciones a la Comisión sobre un «marco de los aspectos éticos de la inteligencia artificial» (Procedimiento: 2020/2012(INL), 20 de octubre de 2020). En esta resolución se propone adaptar la regulación comunitaria que ya existe y, además, incorpora una propuesta de reglamento sobre principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas (Gónzalez, 2020).

En octubre del 2020 el Consejo de la Unión Europea publicó unas conclusiones sobre la aplicación de la CDFUE «en el contexto de la inteligencia artificial y el cambio digital». En la conclusión 16 se establece que la finalidad es el uso de la IA para facilitar la labor de las fuerzas y cuerpos de seguridad y la búsqueda de pruebas fiables en causas penales.

La Agencia de Derechos Fundamentales de la UE tiene como fin institucional asesorar en materia de derechos fundamentales a las Instituciones de la UE y, en diciembre del 2020, la Agencia publicó un estudio sobre la IA y los derechos fundamentales (Consejo de la UE, 2020). El estudio lo que hace es seleccionar algunos de los usos actuales de la IA para poder así identificar qué derechos fundamentales se están vulnerando o están en riesgo, con ello, se proponen medidas concretas para minimizar los riesgos identificados. El principal riesgo identificado es el de vulneraciones del derecho a la intimidad, a la igualdad y a la tutela judicial efectiva.

Pese a todos estos esfuerzos, la normativa respecto a la tipificación de las *deepfakes* de contenido sexual es bastante laxa a nivel europeo, encontrándonos que solo el Reglamento sobre Inteligencia Artificial aborda dicha cuestión. Aunque, podemos encontrar ciertas notas en diferentes normativas que intentan tipificar dichos delitos. Como, por ejemplo, en la Propuesta de Directiva sobre la lucha contra la violencia contra las mujeres y la violencia doméstica (COM/2022/105, 8 de marzo de 2022) se ha recogido, en el considerando 19 y en el artículo 7[1], la necesidad de tipificar la producción, manipulación o difusión no consentida de material íntimo o manipulado. Se ha incluido expresamente la alusión a la edición o fabricación de *deepfakes*.

En este mismo sentido, se han recogido enmiendas a la Propuesta de Reglamento por el que se establecen normas para prevenir y combatir el abuso sexual de los menores (Comisión de Derechos de las Mujeres e Igualdad de género a la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas para prevenir y combatir el abuso sexual de los menores, 8 de mayo de 2023).

3.2 Normativa Española

A nivel nacional, también se han producido esfuerzos legislativos para regular la IA. El primer caso lo encontramos en la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos (en adelante Ley 11/2007), en cuya exposición de motivos podemos extraer que se creó para hacer frente a la evolución de las comunicaciones electrónicas y, poner a disposición de la ciudadanía una Administración actualizada y competente en materia tecnológica.

A partir de este punto, se ha continuado realizando un seguido de normativas que intentan modernizar el proceso con el uso de las TIC. Un ejemplo de ello es la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno, así como la Ley 39/2015, de 1 de octubre, del procedimiento Administrativo Común de las Administraciones Públicas y la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público.

[1] El artículo 7 de la Propuesta de Directiva sobre la lucha contra la violencia contra las mujeres y la violencia doméstica (COM/2022/105, 8 de marzo de 2022), establece: "Los Estados miembros se asegurarán de que se castiguen como infracción penal las siguientes conductas intencionadas: a)divulgar a una multitud de usuarios finales, mediante el uso de las tecnologías de la información y de las comunicaciones, imágenes, vídeos u otros materiales íntimos que representen actividades sexuales de otra persona sin el consentimiento de esta; b)producir o manipular y, posteriormente, divulgar a una multitud de usuarios finales, mediante el uso de las tecnologías de la información y de las comunicaciones, imágenes, vídeos u otros materiales, haciendo que parezca que otra persona está realizando actividades sexuales, sin el consentimiento de esta; c)amenazar con cometer las conductas mencionadas en las letras a) y b) con el fin de coaccionar a otra persona para que realice o acceda a que se realice determinado acto o se abstenga de realizarlo".

En cuanto a la regulación de las IA no es hasta la Ley 15/2022, de 12 de julio, sobre integral para la igualdad de trato y la no discriminación, que se introduce esta primera regulación. A pesar de no ser una Ley que verse íntegramente sobre la IA, en su artículo 23 se hace hincapié sobre el tema, estableciendo que regula derechos y obligaciones de las personas, físicas o jurídicas, públicas o privadas, establece principios de actuación de los poderes públicos y prevé medidas destinadas a prevenir, eliminar, y corregir toda forma de discriminación, directa o indirecta, en los sectores público y privado.

Finalmente, la última regulación importante en nuestro país ha sido la creación el 22 de agosto de 2023 el Real Decreto 729/2023, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial. En su Disposición adicional centésima trigésima, se establece la creación de la Agencia Española de Supervisión de Inteligencia Artificial. Dicha Agencia tiene como objetivo asumir todas las competencias en materia de inteligencia artificial que corresponden a España como Estado miembro de la UE. Es decir, tal y como se establece en dicha disposición, se encarga de supervisar la aplicación y ejecución de lo dispuesto en la mencionada Ley de Inteligencia Artificial, así como de coordinar las actividades encomendadas a los Estados miembros, actuar como el punto de contacto único para la Comisión, y representar al Estado miembro ante el Comité Europeo de Inteligencia Artificial.

Es necesario poner en relieve el artículo 18.4 CE: “La ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos”, siendo este el único precepto que hace referencia al uso de las nuevas tecnologías y su posible incidencia en los derechos fundamentales.

Por todo ello, podemos concluir que la ausencia de marcos normativos supone un importante reto y esfuerzo interpretativo para los juristas al objeto de determinar cómo aplicar el derecho vigente a los sistemas de IA y constituye una preocupación constante para los expertos y para el propio legislador (Asaro, 2007).

3.2.1 Posible tipificación en el Código Penal Español

Tal y como se ha establecido, en el Derecho Penal Español, no existe un marco normativo concreto que tipifique las *deepfakes* pornograficas, es por ello que a la hora de analizar la posible responsabilidad penal derivada de actos u omisiones llevados a cabo por sistemas inteligentes y su posible imputación a los mismos, exige partir necesariamente del elemento de culpabilidad, esto es, no hay pena sin dolo o imprudencia —nullum crimen sine culpa (Muñoz, 2022). Esta cuestión ha sido abordada por autores como Domínguez Peco (2018) o Mengotti (2016) que evidencian la imposibilidad de imputar responsabilidad penal a un robot

o sistema dotado de IA, sin perjuicio de que pueda ser un instrumento o medio para la comisión de un delito.

De conformidad con lo previsto en el artículo 5 Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal, no puede haber pena sin dolo o imprudencia. Dicho precepto enlaza con lo dispuesto en sus artículos 27 y 28 del mismo que establecen que serán responsables criminalmente de los delitos, los autores y los cómplices, y se considerarán autores de un delito a quienes realicen el hecho por sí solos, conjuntamente o por medio de otro del que se sirven como instrumento. Asimismo, conforme a dicho precepto, se considerarán también autores los que induzcan directamente a otro u otros a ejecutarlo, así como los que cooperen a su ejecución con un acto sin el cual no se habría efectuado.

Conforme al ordenamiento jurídico vigente en España no es posible imputar responsabilidad penal a un sistema de IA, sin perjuicio de su utilización como medio o instrumento para su comisión (Muñoz, 2022). Por lo que, únicamente se tendrán en cuenta cuando los sistemas de IA sean utilizados como instrumento para la comisión de un delito por parte de un sujeto, sobre el que recaerá la imputación de la responsabilidad penal y sus consecuencias, en base al dolo o imprudencia concurrente[2].

3.2.1.1 Tipificación de las *deepfakes* sexuales en las que la víctima es menor de edad

En cuanto a los tipos penales que permitirían la persecución de estas prácticas, a través del artículo 189, artículo que tipifica el delito de pornografía infantil, estableciendo que será castigado con la pena de prisión de uno a tres años:

- a) El que utilizare a menores de edad o a incapaces con fines o en espectáculos exhibicionistas o pornográficos, tanto públicos como privados, o para elaborar cualquier clase de material pornográfico, o financiare cualquiera de estas actividades.

- b) El que produjere, vendiere, distribuyere, exhibiere o facilitare la producción, venta, difusión o exhibición por cualquier medio de material pornográfico en cuya elaboración hayan sido utilizados menores de edad o incapaces, aunque el material tuviere su origen en el extranjero o fuere desconocido.

Por lo que, se puede sustraer del artículo que se entiende por pornografía infantil cualquier imagen en la que menores aparezcan en una conducta sexual, pudiendo ser las mismas reales

[2] El hecho de utilizar plataformas como medio de difusión plantea problemas de decomiso o cierre de las páginas web, la responsabilidad penal en cascada de los servidores (artículo 30 CP).

o simuladas, pero en todo caso debiendo ser "realistas". Siguiendo esta línea, en la Circular 2/2015 de la Fiscalía General del Estado, Doctrina de la Fiscalía General del Estado, sobre los delitos de pornografía infantil tras la reforma operada por Ley Orgánica 1/2015, de 19 de junio, establece que las imágenes generadas por ordenador son igualmente sancionables a través del artículo 189 del código penal, siempre y cuando sean realistas. No serían delito de pornografía infantil virtual, por tanto, aquellas imágenes que no cumpliesen con el requisito de ser realistas.

Tal y como se expone en el apartado 2.2 de dicha Circular, la STS nº 271/2012, de 26 de marzo siguiendo al Consejo de Europa, ha definido la pornografía infantil como cualquier material audiovisual que utiliza niños en un contexto sexual. En todo caso, es necesario que el material visual se centre bien en un comportamiento sexual de un menor o bien en sus órganos sexuales. Siguiendo esta misma línea, la STS nº 1342/2003, de 20 de octubre, considera que la imagen de un desnudo no puede ser considerada objetivamente material pornográfico. Por el contrario, el desnudo con connotaciones sexuales sí puede integrarlo, así, para el ATS nº 521/2013, de 21 de febrero "las fotos realizadas por el acusado a la menor pueden considerarse como pornográficas ya que muestran la zona púbica de la niña, su imagen desnuda y del busto en actitud sugerente".

El apartado 2.3 de dicha circular se centra en analizar la denominada pornografía virtual, entendiendo que esta hace referencia a aquella en la que la imagen del menor es una creación artificial pero realista, elaborada por ordenador u otro medio. Establece que el nuevo art. 189.1.d, tras la reforma 1/2015, da cumplimiento al mandato del art. 5 de la Directiva de 2011 y tipifica las conductas relativas a los materiales virtuales. Se entiende como tal en el nuevo precepto, asumiendo literalmente la definición de la Directiva, las imágenes realistas de un menor participando en una conducta sexualmente explícita o imágenes realistas de los órganos sexuales de un menor, con fines principalmente sexuales (art. 189.1.d).

De acuerdo con esta nueva redacción, se entiende que dichas imágenes deben ser realistas, que de acuerdo con el Diccionario de la Real Academia de la Lengua Española "realista" significa que "trata de ajustarse a la realidad". Por tanto, "imágenes realistas" serán imágenes cercanas a la realidad, a la que tratan de imitar, es decir, imágenes que no son reales, pero lo parecen.

En su apartado 2.4 se analiza la pornografía técnica, que puede definirse como aquel material que se integra por imágenes en las que aparecen personas presentadas como menores en un contexto sexual explícita, real o simulada, o cualquier representación de los órganos sexuales de una persona que parezca ser un menor, con fines principalmente sexuales, salvo que la persona que parezca ser un menor resulte tener en realidad dieciocho años o más en el momento de obtenerse las imágenes (art. 189.1.c).

La clave estará, pues, en determinar si los protagonistas del material pueden ser menores y se les presenta como menores. Si pueden ser menores y se les presenta como menores, el material será subsumible en el concepto de pornografía infantil, salvo que se acredite que eran adultos. Si pueden no ser menores y no se les presenta como menores, habrá de optarse por la no incriminación.

Por ende, la Fiscalía se ampara, resumidamente, en que el informe del Consejo Fiscal de 8 de enero de 2013, la Directiva 2011/93/UE o la exposición de motivos de la Ley Orgánica 11/1999, de 30 de abril, permitirían entender que el bien jurídico protegido en el art. 189 va más allá de la indemnidad sexual, llegando a proteger también la dignidad (individual o de la infancia) o la integridad psíquica (Diario Red, 2023).

3.2.1.2 Tipificación de las *deepfakes* sexuales en las que la víctima es mayor de edad

En el caso de que la víctima fuese mayor de edad, no existe ningún delito actualmente que castigue estas conductas de manera específica. Aún así, este se podría intentar aplicar en el delito de distribución o grabación del artículo 197.7 o el delito comodín por excelencia, que no es otro que el delito contra la integridad moral (art. 173), así como un delito de injurias (art 208). Aunque cabe destacar, que se podrían perseguir acudiendo a la vía civil por vulneración del derecho al honor, la intimidad y la propia imagen.

Entrando al analizar el primero de ellos, de acuerdo con el artículo 197 del Código Penal, el cual contiene la regulación del delito de descubrimiento de secretos, estableciendo que será castigado con una pena de prisión de tres meses a un año o multa de seis a doce meses el que, sin autorización de la persona afectada, difunda, revele o ceda a terceros imágenes o grabaciones audiovisuales de aquélla que hubiera obtenido con su anuencia en un domicilio o en cualquier otro lugar fuera del alcance de la mirada de terceros, cuando la divulgación menoscabe gravemente la intimidad personal de esa persona.

Conforme con la Circular 3/2017, de 21 de septiembre, se establece que en el apartado séptimo del artículo 197 se tipifican aquellos delitos en los sin autorización de la persona afectada, difunda, revele o ceda a terceros imágenes o grabaciones audiovisuales de aquélla que hubiera obtenido con su anuencia en un domicilio o en cualquier otro lugar fuera del alcance de la mirada de terceros, cuando la divulgación menoscabe gravemente la intimidad personal de esa persona. Para que el precepto sea aplicable es necesario que la grabación se haya llevado en un marco espacial de carácter reservado, es decir, que se haya obtenido en un domicilio, o en un lugar fuera del alcance de la mirada de terceros, y con consentimiento o anuencia del afectado por ello. La interpretación del concepto de domicilio, no ofrece

dificultad alguna y ha sido objeto de manera reiterada de interpretación jurisprudencial, entre ellas, la STS nº 731/2013 de 7 de octubre, establece que este concepto ha de entenderse de modo amplio y flexible ya que trata de defender los ámbitos en que se desarrolla la vida privada de la persona, debiendo interpretarse a la luz de los principios que tienden a extender al máximo la protección a la dignidad, a la intimidad de la persona y al desarrollo de su privacidad.

La conducta típica, por lo tanto, consiste en difundir, revelar o ceder a terceros las referidas imágenes sin la autorización de la persona afectada, aunque no será necesario para ello acreditar una negativa expresa, sino que podrá ser bastante con la no constancia de autorización.

Existe mayor complejidad en la interpretación de la expresión "lugar fuera del alcance de la mirada de terceros". Dentro de dicho concepto, se incluye cualquier local comercial no abierto al público, o también un lugar al aire libre, si bien en este caso habría que acreditar que reúne garantías suficientes de privacidad de tal forma que pueda asegurarse que las escenas/imágenes, captadas o grabadas, lo fueron en un contexto de estricta intimidad y sustraído a la percepción de terceros ajenos a ellas. En ese sentido el concepto tercero habría que entenderlo referido a personas ajenas al acto o situación objeto de grabación.

Lo que el Legislador pretende es dejar constancia de que las imágenes que posteriormente se difunden tenían, en su origen, un carácter estrictamente privado. Es por ello, que en este precepto podría tener cabida el hecho de generar *deepfakes*, aunque sería necesaria una reforma del artículo, ya que el Legislador define esta situación de privacidad o intimidad de una manera excesivamente cerrada.

El problema radica en que en los delitos de *deepfakes* sexuales, se trata de imágenes falsas, es decir, "fakes", que se obtienen de las redes sociales de la persona. Y, tal y como establece el AAP Cádiz 445/2016, de 10 de noviembre, es difícil hablar de intimidad «en relación con episodios o incidencias (grabación de vídeo) que han sido aireadas, difundidas y divulgadas por quien invoca esa "confidencialidad", pues como tal debemos entender quien lo sube a un grupo de 27 personas». La sentencia parte de la consideración del vídeo como dato inicialmente comprendido por el 197.2 (proceder que se ha criticado supra), y concluye que las circunstancias mencionadas no permiten calificarlo de «reservado»; concluyendo que no existe menoscabo grave de la intimidad.

De acuerdo con la Circular, en cualquier caso, pueden concurrir las diferentes formas de participación que contemplan los artículos 28 y 29 CP. Así, cabría la coautoría cuando dos o más personas comparten el dominio del hecho y obtienen las imágenes que posteriormente y

sin autorización distribuyen, y la cooperación necesaria y la inducción en quienes, sin haber intervenido en la obtención de la imagen y antes de la inicial transmisión, inducen o cooperan con los autores en la divulgación o cesión de los contenidos a otras personas. Es por ello, que la persona afectada, puede utilizar dicho precepto para la persecución de tal circunstancia.

Cuestión distinta es la actuación de los terceros que, sin haber intervenido en la acción inicial, reciben en un momento los contenidos comprometidos y los transmiten. Dichos comportamientos, en principio, únicamente podrían dar lugar a la utilización de los mecanismos previstos en la L.O 1/1982 de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen.

Por todo ello, este artículo, no se podría utilizar para perseguir los delitos cometidos por *deepfakes* sexuales, ya que, de acuerdo a la STC 134/1999 (RTC 1999, 134), FJ 5, el derecho a la intimidad garantiza que a nadie se le puede exigir que soporte pasivamente la revelación de datos, reales o supuestos, de su vida privada personal o familiar. O como también se dijo en la citada STC 176/2013 (RTC 2013, 176) , FJ 7, lo que el art. 18.1 CE garantiza el secreto sobre nuestra propia esfera de vida personal y, por tanto, veda que sean los terceros, particulares o poderes públicos, quienes decidan cuáles son los contornos de nuestra vida privada. Por ello, como se trata de imágenes falsas, es decir, "fakes", no pueden afectar a la intimidad de la persona, ya que, revelan hechos que nunca sucedieron.

Por otro lado, gran parte de los autores y abogados penalistas, como Noelia Bauzá, establecen que dicho delito podría subsumirse en la conducta que tipifica el artículo 173.1 del Código Penal. Este tipifica la conducta en la que una persona inflige a otra persona un trato degradante, menoscabando gravemente su integridad moral. De acuerdo con la STS n.º 98/2020, de 14 de abril, se entiende por trato degradante aquellos actos que pueden "crear en las víctimas sentimientos de temor, de angustia y de inferioridad, susceptibles de humillarles, de envilecerles y de quebrantar en su caso su resistencia física y moral".

En cuanto a los elementos que conforman el concepto de atentado contra la integridad moral, de acuerdo con la STS n.º 181/2023, de 15 de marzo, se establece que estos son:

A. Un acto de claro e inequívoco contenido vejatorio para el sujeto pasivo.

B. La concurrencia de un padecimiento físico o psíquico.

C. Que el comportamiento sea degradante o humillante con especial incidencia en el concepto de dignidad de la persona-víctima.

Además, la jurisprudencia ha establecido unas bases que nos orientan a la hora de identificar si nos encontramos ante este tipo de delito. Por ejemplo, en la STS núm. 233/2009 de 3 de marzo o en la STS núm. 957/2007 de 28 de noviembre, establecen que:" a) Un acto de clara e inequívoco contenido vejatorio para el sujeto pasivo b) un padecimiento, físico o psíquico de dicho sujeto y c) un comportamiento que sea degradante o humillante e incida en el concepto de dignidad de la persona afectada por el delito".

Aún así, para poder condenar a una persona por el 173.1 CP es necesario que concurra, además de esas notas de degradación, humillación y vejatoriedad, que ese trato degradante menoscabara gravemente la integridad moral de la víctima, de este modo sólo podrá castigarse cuando el trato sea especialmente lesivo (STS n.º 98/2020, de 14 de abril).

La sentencia del TS del 3 de marzo de 2009, conocida por su gran labor de interpretación del art. 173.1 CP, desgrana el significado de degradar, estableciendo que se entiende como tal el "privar a alguien de las dignidades, honores, empleos y privilegios que tiene" relacionándolo con una serie de preceptos como el artículo 15 de la Constitución Española, el artículo 5 de la Declaración Universal de Derechos Humanos, etc. concluyendo que el objetivo es "infligir un sufrimiento físico-psíquico tendiente a humillar a la víctima ante los demás o ante sí mismo". Con esta definición, nos esclarece que no es necesario que intervengan o estén presentes terceras personas, sino que es suficiente que la víctima se sienta humillada. Tal y como se establece en la SAP Burgos 109/2015, 27 de marzo de 2015 se considera conductas o tratos degradantes como: "desnudar a la víctima de cintura para abajo durante dos horas (sentencia nº. 1122/98 de 29 de septiembre); desnudar a una persona y obligarla a hacer flexiones n(sentencia nº. 457/03 de 14 de Noviembre); pintar con un spray rosa a una persona que previamente habían desnudado (sentencia nº. 819/02 de 8 mayo), etc". Por lo tanto, es necesario que exista vínculo físico, es decir, que para considerarse trato degradante es imperativo que se realice en presencia del autor o con intervención de éste.

Además, esta sentencia también realza la problemática que trae consigo la interpretación del término "gravemente" del propio art. 173.1 CP y recuerda que son los mismos tribunales, los que valorarán caso a caso si dicha conducta produce un menoscabo grave. El Tribunal Supremo en la sentencia 1023/2021, de 17 de enero de 2022 establece que: "en efecto, el núcleo de la descripción típica está integrado por la expresión "trato degradante" que —en cierta opinión doctrinal— parece presuponer una cierta permanencia, o al menos repetición, del comportamiento degradante, pues en otro caso no habría «trato» sino simplemente ataque; siempre que en ella se aprecie una intensidad lesiva para la dignidad humana suficiente para su encuadre en el precepto; es decir, un solo acto, si se prueba brutal, cruel o humillante puede ser calificado de degradante si tiene intensidad suficiente para ello".

La Sentencia TS, núm. 351/2021, de 28 abril, establece que: “no tiene que ver con el número de actos, sino con la creación de un estado permanente de violencia que afecta a la estructura básica de la convivencia desde el respeto y la dignidad de la persona”. Es por ello, que en el caso de las *deepfakes* se deberá estudiar si se entiende que este tiene suficiente entidad como para considerarse degradante. Existen pocos estudios que hayan explorado los daños que generan las *deepfakes* pornograficas, pero, aún así, existe una gran cantidad de evidencias que sugieren que las víctimas tienen más probabilidades de generar consecuencias adversas, como el trastorno de estrés postraumático. ansiedad, depresión, etc. Un ejemplo de ello es el caso de Laura Escanes, la cual recibió un mensaje en el que le mostraban una serie de fotos de ella desnuda, a raíz de esto, decide darse de baja de manera temporal de las redes, ya que, tal y como ella lo comunica a través de su Twitter, se sentía utilizada (20minutos, 2023).

Finalmente, dichos delitos cometidos a través de *deepfakes* pornograficas se podrían subsumir en el delito del artículo 208 CP, es decir, un delito de injurias. Se entiende por injuria toda acción o expresión que lesiona la dignidad de otra persona, menoscabando su fama o atentando contra su propia estimación.

En el delito de injurias, dispone la STS n.º 344/2020, de 25 de junio: "(...) el bien jurídico protegido es el derecho al honor, que no solo se conforma con la fama que pueda tener una persona, esto es, con su valoración social o con la consideración que de ella puedan tener terceras personas, sino que comporta también que nadie puede ser despreciado en su respeto personal más elemental, impidiendo que pueda sufrir una sensación de bajeza humana que pueda socavar la propia autoestima del individuo, ya que el derecho constitucional al honor (artículo 18 CE) tiene por fundamento la dignidad humana. De este modo, pueden ser constitutivas de delito las injurias las acciones o expresiones dirigidas a menoscabar la dignidad de una persona por más que no se desvele públicamente la persona contra la que se dirigen, siempre que el sujeto pasivo del delito las perciba y que objetivamente sean adecuadas para degradar o menoscabar su consideración como persona".

Para la existencia de un delito de injurias, se requiere la concurrencia de tres elementos fundamentales (sentencia de la Audiencia Provincial de Burgos n.º 270/2015, de 15 de junio):

- Elemento objetivo: constituido por los actos o las expresiones proferidas respecto de las que el sujeto pasivo se sintió atacado, menospreciado o desacreditado.

- Elemento subjetivo: supone la intención, como dolo específico de causar y originar el perjuicio antes señalado.

- Elemento circunstancial: aglutina cuantos factores o datos personales, de ocasión, lugar, tiempo, forma, etc., valorativamente apreciados, contribuyan, de una parte, a esclarecer la verdadera intención o propósito que animaba al sujeto que profiera la ofensa y, de otra, contribuyan a determinar la importancia y magnitud de los tipos del Código Penal.

Por derecho al honor debemos entender que nos referimos tanto a la dignidad que uno tiene de sí mismo como, a la reputación o prestigio que uno tiene frente al resto de la sociedad; bien sea en el ámbito personal o incluso en el ámbito profesional.

Por lo tanto, la conducta típica presenta dos modalidades de ejecución: la acción y la expresión. Mediante una u otra modalidad, se pueden imputar como injurias, tanto hechos como los juicios de valor atentatorios a la dignidad de otra persona. Pero las injurias consistentes en la imputación de hechos no se considerarán graves salvo cuando se hayan llevado a cabo con conocimiento de su falsedad o temerario desprecio hacia la verdad, previsión ésta que confiere un amplio espectro al ejercicio de la libertad de información constitucionalmente tutelada.

Cabe destacar, que en el artículo 209 del Código Penal se castiga con una mayor pena las injurias graves hechas con publicidad. En estos casos, de acuerdo con la sentencia del Tribunal Supremo n.º 344/2020, de 25 de junio, lo que se hace es agredir la autoestima del sujeto pasivo, potenciándose o multiplicándose la lesividad de los hechos mediante instrumentos de divulgación pública que fortalezcan la acción expresamente emprendida para atacar el bien jurídico. En el artículo 211 del Código Penal se define qué se entiende hecha por publicidad cuando se propaguen por medio de la imprenta, la radiodifusión o por cualquier otro medio de eficacia semejante. La fórmula empleada por el legislador permite incluir medios o instrumentos de difusión varios, ya que es imposible enumerarlos todos.

Por lo tanto, este precepto, parece el más oportuno a la hora de tipificar las *deepfakes* de contenido sexual, ya que se trata de una recreación mediante tecnología *deepfake* con la intención de menoscabar el honor, fama, dignidad o estimación de una persona, de las que se tiene conocimiento de su falsedad o temerario desprecio hacia la verdad y, además, se está difundiendo públicamente sin la autorización necesaria.

Pese a que el Grupo Parlamentario Plurinacional SUMAR presentó una Proposición de Ley Orgánica de regulación de las simulaciones de imágenes y voces de personas generadas por medio de la inteligencia artificial, en este, se crea un nuevo artículo, el artículo 208 bis del Código Penal para dar entrada a las acciones injuriosas en las que para su comisión se utilicen simulaciones de imágenes o voces de personas generadas por inteligencia artificial, en tanto

se debe entender la generación y difusión de imágenes o voces artificiales a través de la inteligencia artificial como de alto riesgo potencial de incurrir en conductas injuriosas con relevancia penal.

3.3 Derecho Comparado

La inteligencia artificial nos enfrenta a nuevos campos de infracción de la ley que no existían hace apenas unos años (Bates, 2016). En España, por ejemplo, no existe jurisprudencia sobre delitos cometidos a través de la inteligencia artificial. Aún así, no son pocos los casos de *deepfakes* de contenido sexual en nuestro país, uno de los últimos casos es el de los menores de Almendralejo, concretamente, se trata de un caso de pseudopornografia. En este caso, tal y como hemos adelantado anteriormente, este delito quedaría encuadrado en el artículo 189 del Código Penal.

En el caso de personas mayores de edad, entramos en un tema sin precedentes, aún no ha llegado al Tribunal Supremo ningún caso por videos sexuales no consentidos creados con Inteligencia Artificial, por lo que no hay jurisprudencia. Encontramos ejemplos tanto en personas famosas, que denunciaron la publicación de *deepfakes* de contenido sexual como de personas anónimas, entre las que se encuentran, por ejemplo, Laura Escanes y Rosalia. De acuerdo con el Dictamen 393/2021, del Consejo de Estado de 10 de junio, sobre el Anteproyecto de Ley Orgánica de Ley Integral de Libertad Sexual, se establece que para que dicho delito pueda tipificarse vía penal, debe existir un umbral mínimo de alteración de la vida cotidiana de la víctima, en aplicación del principio de intervención mínima del Derecho Penal, que en nuestro sistema constitucional debe ser una última ratio por su seguridad.

En relación con otros ordenamientos, parece interesante analizar como Estados Unidos tipifica este tipo de delitos, dado que allí se documentaron los primeros casos de *deepfakes* sexuales.

El sistema legal estadounidense tiene dos niveles, el estatal y el federal. Mientras que existe una tipificación de los delitos de *deepfakes* sexuales en las que se presentan menores de edad, tanto a nivel estatal como federal, no existe una regulación común respecto a estas cuando las víctimas son adultos.

La ley federal prohíbe la posesión, distribución y creación de pornografía infantil. La definición de pornografía infantil incluye cualquier representación visual de un menor participando en una conducta sexualmente explícita, incluidas imágenes generadas por ordenador en las que no se puede distinguir si es un menor real o han sido creadas, adaptadas

o modificadas para representar a un menor real identificable (18 U.S.C §§ 2252, 2252A, 2256(8) (Pirus, 2024)).

Las leyes estatales, también prohíben la creación, distribución y posesión de pornografía infantil. Muchas de estas leyes, lo que hacen es beber de la ley federal, por ejemplo, la ley de Virginia define la "pornografía infantil" como "material visual sexualmente explícito que involucra a un menor identificable, incluidas representaciones visuales adaptadas o modificadas" (Fla. Stat. § 827.071). Aí como, las leyes de Minnesota que prohíben la posesión y creación de pornografía infantil incluyen específicamente una "imagen o fotografía generada por ordenador que ha sido creada, adaptada o modificada para que parezca que un menor identificable está participando en una conducta sexual" (Minn. Stat. §§ 617.246 (Pirus, 2024)).

En el ámbito de *deepfakes* de contenido sexual en que la víctima es adulta, la ley federal no cuenta con ninguna regulación específica. En cambio, a nivel estatal, varios estados cuentan con una regulación específica en las que se prohíbe la creación, posesión o distribución de *deepfakes* pornograficas. Cabe destacar que las leyes varían considerablemente respecto a los diferentes estados (Pirus, 2024).

Florida tipifica como delito grave de tercer grado publicar, es decir, el publicar o compartir intencional y maliciosamente una *deepfake* sexual sin consentimiento, entendiendo que como deepfake cualquier representación visual en la que se modifica, altera o adapta la imagen o video para representar una versión realista de una persona identificable. (Estatuto de Florida § 836.13 (Pirus, 2024).)

Hawái considera una invasión de la privacidad de primer grado crear, revelar o amenazar con revelar *deepfakes* que representen a una persona participando en una conducta sexual. Para ser condenado, el fiscal debe demostrar que el acusado tenía la intención de dañar sustancialmente la salud, la seguridad, la carrera, el negocio, etc., de la persona representada (Haw. Rev. Stat. § 711-1110.9 (Pirus, 2024).)

La ley de Utah prohíbe la distribución ilegal de las *deepfakes*, definida como cualquier representación visual o imagen generada por ordenador creada, editada, manipulada o alterada para representar la imagen de un individuo identificable. Es necesario que el sujeto activo tenga conocimiento del daño emocional o angustia que pueda causar a la persona representada (Utah Code Section 76-5B-205).

Nueva York tipifica el porno deepfake a través de sus leyes de pornografía de venganza. El estado amplió su ley actual para prohibir la distribución no consensuada de una imagen

sexualmente explícita de otra persona, incluidas imágenes creadas o alteradas mediante digitalización. Para una condena, el fiscal debe demostrar que el acusado tenía la intención de dañar el bienestar emocional, financiero o físico de la persona representada. (Código Penal de Nueva York § 245.15 (Pirus, 2024).)

En Reino Unido un informe titulado "Destrozando vidas y mitos: un informe sobre Abuso Sexual", dirigido por la profesora Clare McGlynn de la Universidad de Durham, sirvió como una llamada de atención para la comunidad jurídica del Reino Unido, para hacer frente al creciente problema de las *deepfakes* (Hall, Pester & Atanasov, 2022). A esto siguió un comunicado del Ministerio de Justicia que se publicó el 26 de junio de 2019 anunciando la creación de una Comisión Jurídica que debía examinar si la legislación actual era adecuada para abordar este tipo de delitos. El resultado fue el Proyecto de Ley de Seguridad en Línea (Online Safety Bill). En materia de *deepfakes* de contenido sexual, destaca el artículo número 163 titulado "Envío de fotografía o videos de genitales" en este se define como video una imagen en movimiento, incluyendo en dicha definición aquellas imágenes generadas por ordenador o de cualquier otra manera que parezca una fotografía o video. Dicho precepto, por lo tanto, simplemente tipifica la distribución, pero no la creación de *deepfakes* sexuales.

El proyecto de ley, a diferencia de la legislación estadounidense, no requiere que el perpetrador tenga la intención de causar angustia o humillación. Esto significa que no se pedirá a las víctimas que lo demuestren ante los tribunales (Williams, 2023).

En cuanto a los menores de edad, en Reino Unido, se utiliza la normativa existente para tipificar dichas conductas, como sería la Sección 1 del Acta 1978 (PCA 1978) de la Protección al menor y la Sección 160 del Acta de 1988 (CJA 1988) Justicia Criminal. En las que se condena el hacer, distribuir, la tenencia o publicación de dichas fotografías o videos.

En 2020, el gobierno de Corea del Sur promulgó una revisión de la "Ley sobre casos especiales relacionados con el castigo, etc. de delitos sexuales", está prohíbe tanto la creación como la distribución de *deepfakes* que "puedan provocar deseo sexual o vergüenza contra la persona víctima del video o imagen". Según esta nueva ley, los condenados pueden ser castigados con hasta cinco años de prisión y, en caso de que venda el acceso a las *deepfakes,* la pena se multiplica (Williams, 2023).

Vale la pena señalar que varios factores exclusivos de Corea del Sur pueden haber contribuido a la decisión de adoptar una ley tan restrictiva. Y es que esta política se propuso a raíz de un incidente, el "caso de la enésima habitación", como se lo conoce ahora, que involucró a un grupo de usuarios anónimos de Telegram que atacaron a cientos de mujeres y docenas de menores con horribles actos de violencia sexual, incluidas violaciones y

agresiones físicas, así como la creación de *deepfakes* y otras formas de pornografía no consentida.

Por todo ello, podemos concluir que, pese a no existir una regulación específica sobre la materia, la mayoría de los países, están abordando la situación de manera similar, tipificando la distribución, posesión y publicación.

4. Conclusiones

A lo largo de este TFG se ha expuesto el concepto de lo que se entiende por *deepfakes*, definiendo estas como un producto de la inteligencia artificial que tiene el poder de crear contenidos falsos increíblemente realistas, como imágenes, audio y vídeos. Los *deepfakes* pueden ser tanto herramientas legítimas, como en el ámbito de los videojuegos y el entretenimiento, como también pueden tener un lado más oscuro.

Este TFG tiene como objetivo analizar este ámbito no legítimo de las *deepfakes*, concretamente, intentar esclarecer si los delitos cometidos por *deepfakes* de contenido sexual se podrían subsumir dentro del marco legal y normativo Español. Para ello, la revisión de artículos doctrinales y jurídicos nos ha proporcionado un fundamento teórico sólido para comprender los desafíos que enfrenta el ordenamiento jurídico en la adaptación a la rápida evolución de las tecnologías de manipulación digital.

En concreto, a lo largo del trabajo se ha analizado la normativa europea en materia de *deepfakes*, la cual era inexistente hasta el año pasado en que se elaboró la Ley de Inteligencia Artificial, está, dividirá las *deepfakes* en cuatro categorías principales en función del riesgo potencial que supongan para la sociedad. Las normas generales sobre IA se aplicarán un año después de su entrada en vigor, en mayo de 2025, y las obligaciones para los sistemas de alto riesgo en tres años.

También hemos analizado normativa y legislación de otros países en esta materia, resaltando Estados Unidos, país en el que se originó el fenómeno de las *deepfakes* de contenido sexual a raíz de una publicación en Reddit. Pese a ello, no fue hasta 2023, en que la política de IA en EE. UU. se convirtió en un tema central. Ésto fue a raíz de una orden ejecutiva que exige más transparencia y nuevas normas para la IA que emitió el presidente Biden. Tal y como hemos analizado en el trabajo, no existe una regulación a nivel federal de las *deepfakes* de contenido sexual cuando la víctima es adulta, aunque ya se está empezando a plantear una ley que clasifique los tipos y usos de la IA en función del riesgo que planteen, un marco similar a la Ley de IA de la UE. Pese a la falta de regulación a nivel federal, sí que existen normativas sobre las *deepfakes* sexuales a nivel estatal. Hemos podido constatar que, en la mayoría de los casos, la norma requiere de que exista dolo por parte del sujeto activo y, por lo tanto, que este sea probado. Cosa que entra en contradicción con la legislación y normativa de las *deepfakes* sexuales en Reino Unido, donde independientemente de la intención del sujeto activo, si este publica o envía fotos o vídeos será condenado. Finalmente, la legislación coreana prohíbe tanto la creación como la distribución de *deepfakes* sexuales.

En materia de *deepfakes* de contenido sexual en las que la víctima es menor de edad, en todos los países analizados existe regulación específica. En el caso de Estados Unidos, se prohíbe la tenencia, distribución y creación, en Reino Unido y en España, utilizan la ley existente para tipificar dichos delitos; concretamente, se utiliza en España el artículo 189 CP que tipifica todos los delitos en los que la víctima sea menor de edad. Mientras que en el Reino Unido, se utiliza la Sección 1 del Acta 1978 (PCA 1978) de la Protección al menor y la Sección 160 del Acta de 1988 (CJA 1988) Justicia Criminal.

En España, actualmente, tal y como hemos establecido no existe una regulación específica en materia de *deepfakes* sexuales en adultos. Expertos en la materia entienden que este tipo de delito se podría tipificar a través de los delitos del 197.7 CP de distribución y producción, el 173 CP de la integridad moral y el 208 CP injurias. A lo largo del trabajo se ha analizado si este tipo de *deepfakes* se podrían subsumir en los diferentes tipos, descartando cada uno de ellos por los diferentes problemas que podríamos encontrar en cada uno de ellos. Respecto al 197.2 CP, lo que se pretende proteger es que las imágenes que posteriormente se difunden tenían, en su origen, un carácter estrictamente privado. Pero, tal y como se ha vislumbrado a lo largo del trabajo, dicho precepto no podría utilizarse para subsumir los delitos cometidos por *deepfakes,* ya que, se trata de imágenes falsas, es decir, "fakes", que se obtienen de las redes sociales de la persona. Y, tal y como establece el AAP Cádiz 445/2016, de 10 de noviembre, es difícil hablar de intimidad en relación con grabaciones o vídeos que han sido aireados, difundidos y divulgados por quien invoca dicha "confidencialidad". Además, se trata de imágenes que no vulnerarían el derecho de la intimidad, ya que se trata de imágenes generadas por ordenador que no revelan datos, reales o supuestos, de su vida privada personal o familiar.

En cuanto al artículo 173.1 CP, tal y como se ha analizado, es necesario que concurra un cambio en la vida cotidiana de la persona, es decir, que exista una alteración física o psíquica. Ya que, el objetivo es "infligir un sufrimiento físico psíquico tendiente a humillar a la víctima ante los demás o ante sí mismo". Además, de que exista habitualidad o que sea un único acto que se demuestre ser brutal, cruel o humillante. Pero, para que dicho único acto se considere degradante es necesario que exista vínculo físico, es decir, que para considerarse trato degradante es imperativo que se realice en presencia del autor o con intervención de éste.

Finalmente, tras haber analizado el artículo 208 CP referente al delito de injurias, el cual, pese a que el Código Penal Español no menciona directamente las *deepfakes* de contenido sexual, considero que nuestro ordenamiento, y concretamente en el artículo 208 CP, cumple todos los elementos para poder aplicar el tipo, ya que, se puede aplicar tanto a hechos como a juicios de valor que atenten contra la dignidad. Y, en el caso de tratarse de hechos, se considerarán graves cuando se hayan llevado a cabo con conocimiento de su falsedad o

temerario desprecio hacia la verdad, objetivo de ello es la creación de *deepfakes*, que se define como cualquier obra que pretende engañar a través de estímulos visuales y auditivos.

En resumen, aunque las *deepfakes* de contenido sexual no están específicamente tipificadas en el Código Penal español, existen disposiciones legales que protegen el honor y sancionan la creación y difusión de material no consentido. Es por ello, que es importante seguir debatiendo y actualizando la legislación para abordar los desafíos planteados por la tecnología y la inteligencia artificial.

5. Bibliografía

5.1 Libros

Alzahrani, T., Al-Bander, B., & Al-Nuaimy, W. (2021). Deep Learning Models for Automatic Makeup Detection.

Capistrán, J. B. (2022). Evolution of Deepfake: semantic fields and discursive genres (2017-2021). *Icono14*.

Cerdán Martínez, V., & Padilla Castillo, G. (2019). Historia del fake audiovisual: deepfake y la mujer en un imaginario falsificado y perverso. *Historia y comunicación social*, Universidad Computense de Madrid.

Citron, D. K., & Chesney, R. (2019). Deepfakes: A looming challenge for privacy, democracy and national security. *California Law Review*.

D. Bitouk, N. K. (2008). Face swapping: Automatically replacing faces in photographs. *ACM Digital Library*,

Dominguez, E. (2018). El derecho de los robots. Wolters Kluwer.

Gónzalez, I. (2017). Avances y desafíos en materia de ciberdelincuencia de género a nivel europeo. Editorial Comares.

Mengotti, J. (2016). El derecho penal frente a los robots. Madrid: FIDE Papers.

Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: deeplearning lip sync from audio. *ACM Digital Library*.

5.2 Conferencias

Xingjie, Z., Song, J., & Park, J. (2014). The image blending method for face swapping. *IEEE International Conference Network Infrastructure and Digital Content*, (págs. 95-98).

Zakharov, E., Shysheya, A., Burkov, E., & Lempitsky, V. (2019). Few shot adversarial learning of realistic neutral talking head models. *IEEE International Conference on Computer Vision (CVF)*, (págs. 9459-9468).

5.3 Periódicos y revistas

«Deepfakes» pornográficos: Cuando la IA desnuda tu intimidad y vulnera tus derechos. (2 de noviembre de 2023). *Magisnet*. Consultado el 6 de mayo de 2024.

https://www.magisnet.com/2023/11/deepfakes-pornograficos-cuando-la-ia-desnuda-tu-intimidad-y-vulnera-tus-derechos/

Adjer, H., Patrini, G., & Cavalli, F. (2019). Automating image: deepfake bots on Telegram. *Personal Blog*. Consultado el 6 de mayo de 2024. https://giorgiop.github.io/posts/2020/10/20/automating-image-abuse/

Adriani, R. (2019). The evolution of fake news and the abuse of emerging technologies. *European Journal of Social Sciences*. Consultado el 6 de mayo de 2024. https://revistia.org/index.php/ejss/article/view/4241

Aider, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). The State of deepfakes: landscape, threats and impact. *Deeptrace Labs*. Consultado el 6 de mayo de 2024. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Asaro, P. (2007). Robots and responsability from a legal perspective. *University of Umea News*. Consultado el 6 de mayo de 2024. https://peterasaro.org/writing/ASARO%20Legal%20Perspective.pdf

Ayyub, R. (21 de noviembre de 2018). I was the victim of a deepfake porn plot intented to silence me. *Huffpost*. Consultado el 6 de mayo de 2024. https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316

Bates, S. (2016). Revenge Porn and mental health: a qualitative analysis of the mental health effects of revenge porn on female survivors. *Sage Journals*. Consultado el 6 de mayo de 2024. https://journals.sagepub.com/doi/abs/10.1177/1557085116654565

Breland, A. (15 de marzo de 2019). The bizarre and terrifrying case of the deepfake video that helped bring an African nation to the brik. *Mothe Jones*. Consultado el 6 de mayo de 2024. https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/

Chawla, R. (2019). Deepfakes: How a pervert shook the world. *International Journal of Advance, Research and Development*. Consultado el 6 de mayo de 2024. https://www.semanticscholar.org/paper/Deepfakes-%3A-How-a-pervert-shook-the-world-Chawla/c3b3a6d27dbbfed4df630b39fc0a8a6692b1828a

Citron, D. (2014). Hate Crimes in Cyberspace. *Harvard University Press*.Consultado el 6 de mayo de 2024. https://www.hup.harvard.edu/books/9780674659902

Culture Shock. True, Lies and Technology. (2020). *Initiative*.Consultado el 6 de mayo de 2024. https://www.academia.edu/80624402/Deepfake_evolution_Semantic_Fields_and_Discursive_Genres_2017_2021_Evolución_del_Deepfake_campos_semánticos_y_géneros_discurs

ivos_2017_2021_Evolução_do_Deepfake_campos_semânticos_e_gêneros_discursivos_2017_2021?hb-sb-sw=80759596

Deepfakes sexuales generados por IA: machismo pero... ¿delito? (24 de septiembre de 2023). *Diario Red*. Consultado el 6 de mayo de 2024. https://diariored.canalred.tv/actualidad/deepfakes-sexuales-generados-por-ia-machismo-pero-delito/

Diakopoulos, N., & Johnson, D. (2020). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media and Society*, 1-27.Consultado el 6 de mayo de 2024. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3474183

El Gobierno aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial. (5 de mayo de 2023). *Garrigues Digital*.Consultado el 6 de mayo de 2024. https://www.garrigues.com/es_ES/garrigues-digital/gobierno-aprueba-estatuto-agencia-espanola-supervision-inteligencia-artificial

Fletcher, J. (2018). Deepfakes, artificial Intelligence and some kind of dystopia: The new faces of online post-fact performance. *Theatre Journal*, 455-471. Consultado el 6 de mayo de 2024. https://www.proquest.com/docview/2177217909?sourcetype=Scholarly%20Journals

Gabon. (31 de diciembre de 2018). Discours à la nation du président ali bongo ondimba. *Facebook*. Consultado el 6 de mayo de 2024. https://www.facebook.com/tvgabon24/videos/discours-à-la-nation-du-président-ali-bongo-ondimba/324528215059254/?locale=fr_FR

Gardiner, N. (2019). Facial re-enactment, speech synthesis and the rise of the Deepfake. *Theses*. Edith Cowan University. Consultado el 6 de mayo de 2024. https://ro.ecu.edu.au/theses_hons/1530/

Gonzalez, G. (2020). *Artificial Intelligence and Law Enforcement Impact on Fundamental Rights.* Consultado el 6 de mayo de 2024. *https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2020)656295*

Gónzalez, I. (2023). El uso de la inteligencia artificial generativa en la ciberdelincuencia de género: ante el auge de los deepfakes. *Ius et Scientia*, 157-180.Consultado el 6 de mayo de 2024. https://revistascientificas.us.es/index.php/ies/article/view/24691

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Courville, A. y. (2014). Generative Adversarial Networks. *arXiv*. Consultado el 6 de mayo de 2024. https://arxiv.org/abs/1406.2661

Hao, K. (3 de septiembre de 2020). Deepfakes virales: la final línea que separa el humor del abuso. *MIT Technology Review*. Consultado el 6 de mayo de 2024. https://www.technologyreview.es/s/12589/deepfakes-virales-la-fina-linea-que-separa-el-humor-del-abuso

Hart. (1994). *The Concept of Law.* Oxford University Press.Consultado el 6 de mayo de 2024. https://global.oup.com/academic/product/the-concept-of-law-9780199644704?cc=ro&lang=en&

Home Security Heroes. (2023). State of deepfakes: realities, threats and impact. Consultado el 6 de mayo de 2024. https://www.homesecurityheroes.com/state-of-deepfakes/

Korshunova, I., W. Shi, J. D., & Theis, L. (2017). Fast face-swap using convolutional neural networks. *IEEE International Conference on Computer Vision (ICCV)*, (págs. 3697-3705). Consultado el 6 de mayo de 2024. https://openaccess.thecvf.com/content_ICCV_2017/papers/Korshunova_Fast_Face-Swap_Using_ICCV_2017_paper.pdf

La ministra Irene Montero respalda a Laura Escanes en su denuncia por publicar fotos de ella desnuda, manipuladas con IA. (16 de agosto de 2023). *20minutos.* Consultado el 6 de mayo de 2024. https://www.20minutos.es/noticia/5164880/0/ministra-irene-montero-respalda-laura-escanes-su-denuncia-por-usar-fotos-desnuda-ella-manipuladas/#:~:text=2023%20-%2013%3A38h-,La%20ministra%20Irene%20Montero%20respalda%20a%20Laura%20Escanes%20en%20su,"Ha%20llegado%20el%20momento".

Lee, B. (3 de febrero de 2018). Deepfakes porn has serious consequences. *BBC News*. Consultado el 6 de mayo de 2024. https://www.bbc.com/news/technology-42912529

Lee, D. (10 de mayo de 2019). Deepfake Salvador Dalí takes selfies with museum visitors. *The Verge*. Consultado el 6 de mayo de 2024. https://www.theverge.com/2019/5/10/18540953/salvador-dali-lives-deepfake-museum

Meskys, E., Liaudanskas, A., Kalpokiene, J., & Jurcys, P. Regulating deepfakes: legal and ethical considerations. *Journal of Intellectual Poperty Law and Practice*, 24-31. Consultado el 6 de mayo de 2024. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3497144

Mona Lisa "broght to life" with Deepfake AI. (24 de mayo de 2019). *BBC News*. Consultado el 6 de mayo de 2024. https://www.bbc.com/news/technology-48395521

Muñoz, J. (4 de enero de 2022). Inteligencia Artificial y responsabilidad penal. *Diario la Ley*. Consultado el 6 de mayo de 2024. *https://diariolaley.laleynext.es/Content/DocumentoRelacionado.aspx?params=H4sIAAAAAAAEAC2NQUsEMQyFf429CNK6jOIhl3GOi4gO3jNt6BS6zdqm486_N7obeCSPfMn77lT3mS4CnkOKfH-mgtm0vXDZTzDXTkZwaWDvnr1TPRr00jFP7MEdnv5c2mjGBazhGqiOu07CgvmDGjg3DKat_POGW4ooicuI9fo3hQDTbLUOg7P2xWxUmwLwlSIVIbOmuB5VcuUbYfXrO0YCTe8nZfgB2_ly24xdRK8XKZ__3visfUKhV8xUwi33F3ql243zAAAAWKE*

Paris, B. &. (6 de junio de 2021). Deepfakes and sheao fakes: The manipulation of audio and visual evidence. *Data and Society*. Consultado el 6 de mayo de 2024. https://datasociety.net/library/deepfakes-and-cheap-fakes/

Pirius, R. (2024). Is deepfake pornography illegal? Learn where the law stands when it comes to criminalizing deepfake porn at the federal and state levels and the challenges faced by prosecutors and victims. *Criminal Defense Lawyer*. Consultado el 6 de mayo de 2024. https://www.criminaldefenselawyer.com/resources/is-deepfake-pornography-illegal.html

President's words used to create deepfakes at davos. (24 de enero de 2020). *CBS*. Consultado el 6 de mayo de 2024. https://www.youtube.com/watch?v=4A9LAxhi68I

R.D. (9 de julio de 2020). "Welcome to Chechnya" uses deepfake technology to protect its subjects. *The Economist*. Consultado el 6 de mayo de 2024. https://www.economist.com/prospero/2020/07/09/welcome-to-chechnya-uses-deepfake-technology-to-protect-its-subjects?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=18151738051&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gad_source=1&gclid=CjwKCAjw26KxBhBDEiwAu6KXt9rSxYLJ-OVzDaTrs5qp1Q9X1WnNfiSdjNVvw-FnUFPhPqcxjxEn6xoCh_0QAvD_BwE&gclsrc=aw.ds

Roscoe, J., & Hight, C. M. (2001). Documentary and the Subversion of Factuality. *Manchester University Press*. Consultado el 6 de mayo de 2024. https://books.google.es/books/about/Faking_It.html?id=i55kcByOh70C&redir_esc=y

Rösser, A., Cozzolino, D., Verdoliva, L., & Riess, C. T. (2018). Faceforensics: A large scale video dataset for forgery detection in human faces. *ArXiv*. Consultado el 6 de mayo de 2024. https://arxiv.org/abs/1803.09179

Shaoanlu. (2018). Faceswap-GAN: A denoising Autoencoder + Adversarial losses and attention. Mechanisms for face swapping. *GitHub*. Consultado el 6 de mayo de 2024. https://github.com/shaoanlu/faceswap-GAN

Siekierski, B. J. (2019). Deepfakes: what can be done about synthetic audio and video. *Library of Parlamient.* Consultado el 6 de mayo de 2024. https://lop.parl.ca/sites/PublicWebsite/default/en_CA/ResearchPublications/201911E

Silverman, C. (17 de abril de 2018). How to spot a deepfake like the Barack Obama. *BuzzFeed*. Consultado el 6 de mayo de 2024. https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed

Villasenor, J. (14 de febrero de 2019). Artificial intelligence, deepfakes and the uncertain future of trith. *Brookings*. Consultado el 6 de mayo de 2024. https://www.brookings.edu/articles/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/

Villegas, M. (enero de 2010). La impotencia de un legislador bienintencionado: el futuro delito de acoso inmobiliario. *Noticias jurídicas*. Consultado el 6 de mayo de 2024. https://noticias.juridicas.com/conocimiento/articulos-doctrinales/4512-la-impotencia-de-un-legislador-bienintencionado:-el-futuro-delito-de-acoso-inmobiliario/

Vincent, J. (18 de mayo de 2021). Deepfake dubs could help translate films and TV without losing an actor's original performance. *The Verge*. Consultado el 6 de mayo de 2024. https://www.theverge.com/2021/5/18/22430340/deepfake-dubs-dubbing-film-tv-flawless-startup

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology, Innovation Management Review*. Consultado el 6 de mayo de 2024. https://timreview.ca/sites/default/files/article_PDF/TIMReview_November2019%20-%20D%20-%20Final.pdf

Williams, K. (24 de octubre de 2023). Exploring Legal Approaches to regulating Nonconsensual deepfake pornography. *Tech Policy Press*. Consultado el 6 de mayo de 2024. https://www.techpolicy.press/exploring-legal-approaches-to-regulating-nonconsensual-deepfake-pornography/

5.4 Legislación

Agencia de los Derechos Fundamentales (UE) (2021), Construir correctamente el futuro. La inteligencia artificial y los derechos fundamentales. European Union Agency for Fundamental Rights, Fra Uploads

Comunicación de la Comisión Europea (UE) COM/2018/237, para el Parlamento Europeo, el Consejo Europeo, el Consejo, el Comité Económico y social europeo y al Comité de las

Regiones. Inteligencia Artificial para Europa, DGCOMM Documento 52018DC0237, de 25 abril de 2018. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM%3A2018%3A237%3AFIN

Comunicación de la Comisión Europea (UE) COM/2019/168 final, para el Parlamento Europeo, el Consejo, el Comité económico y social europeo y al Comité de las Regiones. Generar confianza en la inteligencia artificial centrada en el ser humano. DGCOMM Documento 52019DC0168, de 8 de abril de 2019. https://eur-lex.europa.eu/legal-content/es/ALL/?uri=CELEX%3A52019DC0168

Comunicación de la Comisión Europea (UE) COM2018/795 final, para el Parlamento Europeo, el Consejo, el Comité económico y social europeo y al Comité de las Regiones. Plan coordinado sobre la inteligencia artificial. DGCOMM Documento 52018DC0795, de 7 de diciembre de 2018. https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=COM%3A2018%3A795%3AFIN

Consejo de la Unión Europea (21 octubre de 2020) (UE). Inteligencia artificial: la Presidencia presenta unas Conclusiones sobre la garantía del respeto de los derechos fundamentales. Comunicado de prensa https://www.consilium.europa.eu/es/press/press-releases/2020/10/21/artificial-intelligence-presidency-issues-conclusions-on-ensuring-respect-for-fundamental-rights/

España. Circular 2/2015, Doctrina de la Fiscalía General del Estado, sobre los delitos de pornografía infantil tras la reforma operada por Ley Orgánica 1/2015, FIS-C-2015-00002 de 19 de junio. https://www.boe.es/buscar/doc.php?id=FIS-C-2015-00002

España. Circular 3/2017, Doctrina de la Fiscalía General del Estado, sobre la reforma del Código Penal operada por la LO 1/2015, de 30 de marzo, en relación con los delitos de descubrimiento y revelación de secretos y los delitos de daños informáticos, Referencia: FIS-C-2017-00003 de 21 de septiembre. https://www.boe.es/buscar/doc.php?id=FIS-C-2017-00003

España. Constitución Española. Boletín Oficial del Estado, 29 de diciembre de 1978, núm. 311 https://www.boe.es/buscar/act.php?id=BOE-A-1978-31229

España. Dictamen num.393/2021. Anteproyecto de Ley Orgánica para la garantía integral de la libertad sexual. Referencia: 393/2021 de 10 de junio de 2021. https://www.boe.es/buscar/doc.php?id=CE-D-2021-393

España. España. Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal. «BOE» núm. 281, de 24/11/1995. https://www.boe.es/buscar/act.php?id=BOE-A-1995-25444

España. Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos. BOE núm. 150, de 23/06/2007. https://www.boe.es/buscar/act.php?id=BOE-A-2007-12352

España. Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación. BOE núm. 167, de 13/07/2022. https://www.boe.es/buscar/act.php?id=BOE-A-2022-11589

España. Ley 18/2011, de 5 de julio, reguladora del uso de las tecnologías de la información y la comunicación en la Administración de Justicia. BOE núm. 160, de 06/07/2011 https://www.boe.es/buscar/act.php?id=BOE-A-2011-11605

España. Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas. BOE núm. 236, de 02/10/2015. https://www.boe.es/buscar/act.php?id=BOE-A-2015-10565

España. Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público. BOE núm. 236, de 02/10/2015. https://www.boe.es/buscar/act.php?id=BOE-A-2015-10566

España. Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial. BOE núm. 210, de 2 de septiembre de 2023, páginas 122289 a 122316 (28 págs.) https://www.boe.es/diario_boe/txt.php?id=BOE-A-2023-18911

España. Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno. BOE» núm. 295, de 10/12/2013. https://www.boe.es/buscar/act.php?id=BOE-A-2013-12887

Estados Unidos. U.S 2022 Hawaii Revised Statutes. Hawaii Penal Code: 711. Offenses Against Public Order: 711-1110.9 Violation of privacy in the first degree. https://law.justia.com/codes/hawaii/2022/title-37/chapter-711/section-711-1110-9/

Estados Unidos. U.S 2022. Title 76 Utah Criminal Code. Utah Code Section 76-5B-205. https://le.utah.gov/xcode/Title76/Chapter5B/76-5b-S205.html

Estados Unidos. U.S 2023 Florida Statutes (including Special Session C) Statutes & Constitution http://www.leg.state.fl.us/statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=0800-0899/0836/Sections/0836.13.html

Estados Unidos. U.S 2023 Minnesota Statutes Sec. 617.246 MN Statutes. https://www.revisor.mn.gov/statutes/cite/617.246

Estados Unidos. U.S The 2023 Florida Statutes (including Special Session C). Statutes & Constitution http://www.leg.state.fl.us/statutes/index.cfm?App_mode=Display_Statute&URL=0800-0899/0827/Sections/0827.071.html

Estados Unidos. U.S The New York State Senate. Penal (PEN) Chapter 40, Part 3, Title N, Article 245 https://www.nysenate.gov/legislation/laws/PEN/245.15

Estados Unidos. U.S. Code § 2252. Certain activities relating to material involving the sexual exploitation of minors. https://www.law.cornell.edu/uscode/text/18/2252

Estados Unidos. U.S. Code § 2252A. Certain activities relating to material constituting or containing child pornography. https://www.law.cornell.edu/uscode/text/18/2252A

Estados Unidos. U.S. Code § 2256. Definitions for chapter. https://www.law.cornell.edu/uscode/text/18/2256

Unión Europea. Propuesta de Directiva del Parlamento Europeo y del Consejo (UE) COM/2022/105 final sobre la lucha contra la violencia contra las mujeres y la violencia doméstica. DGCOMM Documento 52022PC0105, De 8 de marzo de 2022 https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A52022PC0105

Unión Europea. Propuesta de Reglamento (UE) COM/2021/206 final del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión. COD 2021/0106. Documento 52021PC0206, 21 de abril de 2021. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A52021PC0206

Unión Europea. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) «DOUE» núm. 119 de 4 de mayo de 2016, páginas 1 a 88 https://www.boe.es/buscar/doc.php?id=DOUE-L-2016-80807

Unión Europea. Resolución del Parlamento Europeo (UE), Procedimiento: 2020/2012(INL) con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas. DOUE Documento 52020IP0275, de 20 de octubre de 2020. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A52020IP0275

5.5 Sentencias

Sentencia del Tribunal Constitucional núm. 134/1999 (Sala Primera) de 15 julio de 1999; (RTC 1999\134)

Sentencia del Tribunal Constitucional núm. 176/2013 (Sala Primera) de 21 octubre de 2013; (RTC 2013\176)

Sentencia del Tribunal Supremo núm. 507/2019 (Sala de lo Penal, Sección 1) de 25 de octubre de 2019; (Rec. 1473/2018)

Sentencia Tribunal Supremo núm.616/2022, (Sala de lo Penal, Sec. 1) de 22 de junio de 2022; (Rec. 2686/2020)

Sentencia del Tribunal Supremo núm. 351/2021, (Sala de lo Penal, Sección1ª), de 28 abril de 2021; (RJ 2021\2298).

Sentencia Tribunal Supremo núm.771/2023 (Sala de lo Penal, Sec. 1) de 18 de octubre de 202; (Rec. 6313/2021).

Sentencia Tribunal Supremo, núm. 98/2020, (Sala de lo Penal, Sec. 1) de 14 de abril de 2020; (Rec. 103/2019).

Sentencia Tribunal Supremo, núm. 181/2023, (Sala de lo Penal, Sec.1) de 15 de marzo de 2023 (Rec. 3337/2021).

Sentencia Tribunal Supremo núm. 1061/2009 (Sala de lo Penal, Sección 1ª) de 26 de octubre; (Rec. 10339/2009)

Sentencia Tribunal Supremo núm. 957/2007 (Sala de lo Penal, Sección 1º) FD 4º párrafo. 12º. de 28 de noviembre; (Rec. 896/2007)

Sentencia Tribunal Supremo núm. 233/2009 (Sala de lo Penal, Sección 1ª) de 3 de marzo; (Rec. 1732/2008)

Sentencia Tribunal Supremo núm. 344/2020, (Sala de lo Penal, Sec. 1) de 25 de junio de 2020; (Rec. 2151/2019)

Sentencia del Tribunal Supremo, núm 1023/2021 (Sala de lo Penal, Sec 2), 17 de enero de 2022; (Rec. 407/2020. 17-01-22

Sentencia Audiencia Provincial de Burgos, núm. 109/2015 (Sec.1) de 27 de marzo de 2015; (Rec. 4/2015)

Sentencia Audiencia Provincial de Burgos núm. 270/2015 (Sec. 1), de 15 junio de 2015; (Rec. 129/2015)

5.6 Recursos visuales

Advocacy, F. (26 de noviembre de 2019). *Future Advocacy. Boris Johnson has a message for you*. Obtenido de Youtube.

Comedy, C. 4. (15 de marzo de 2021). *Channel 4 Comedy*. Obtenido de The Queen's Christmas message gets a deepfake makeover.

Face, C. S. (6 de agosto de 2019). *Ctrl Shift Face*. Obtenido de Youtube: htt ps://www.youtube.com/watch?v=VWrhRBb-1Ig

Mechanical. (26 de julio de 2020). *Mechanical*. Obtenido de Youtube: https://bit.ly/3suHg08

Ment, M. (20 de agosto de 2020). *Moe Ment*. Obtenido de Youtube: https://bit.ly/3IAzuaH

Pixellot.tv.T. (22 de diciembre de 2021). Obtenido de Youtube: https: //bit.ly/3IpL9IZ

The Zizi Show. (2020). Obtenido de Youtube: https://zizi.ai/