Máster Universitario en Derecho y Violencia de Género

LAS ULTRAFALSIFICACIONES: UNA NUEVA MANIFESTACIÓN DE VIOLENCIA CONTRA LAS MUJERES

Presentado por

MARÍA MATAS PRÓSPER

Dirigido por

DRA. PAZ LLORIA GARCÍA

Curso académico [2024-25]

VNIVERSITAT DE VALÈNCIA Facultat de Dret

ÍNDICE

INTRODUCCIÓN 5

CAPÍTULO I. SOCIEDAD TECNOLÓGICA E INTELIGENCIA ARTIFICIAL 9

1. INTRODUCCIÓN 9

2. IA: DESARROLLO HISTÓRICO Y CONCEPTO 12

3. LAS *DEEPFAKES* O ULTRAFALSIFICACIONES 16

3.1. Concepto, evolución, usos y funcionamiento 16

3.2. Riesgos generales derivados de las *deepfakes* 19

3.2.1. Impacto de las deepfakes en el ámbito político y en la difusión de desinformación 20

3.2.2. Las *deepfakes* como instrumento para cometer delitos económicos 23

3.2.3. Utilización de las *deepfakes* en la manipulación de pruebas en procesos judiciales 25

3.2.4. Empleo de las *deepfakes* para la creación de pornografía. Especial referencia a la pornografía infantil (remisión) 26

4. MÉTODOS DE DETECCIÓN DE LAS *DEEPFAKES* 28

5. REGULACIÓN DE LAS *DEEPFAKES* EN EL ÁMBITO COMUNITARIO Y NACIONAL 30

CAPÍTULO II. LAS *DEEPFAKES* SEXUALES COMO NUEVA FORMA DE VIOLENCIA CONTRA LAS MUJERES 35

1. INTRODUCCIÓN Y CONCEPTO 35

2. LAS *DEEPFAKES* SEXUALES Y OTROS CIBERDELITOS DE GÉNERO: ANÁLISIS Y PERSPECTIVA 39

3. LAS ULTRAFALSIFICACIONES COMO NUEVO MÉTODO DE EJERCER CONTROL Y DOMINACIÓN SOBRE LA MUJER 44

3.1. Las ultrafalsificaciones como herramienta para cometer violencia contra las mujeres en política. Influencia de la manosfera en la creación y difusión de deepfakes de contenido sexual. Especial referencia al término *gendertrolling* 45

3.2. Las *deepfakes* como instrumento para cometer chantaje emocional contra las mujeres. Especial referencia a la pornografía de la venganza y la sextorsión 53

CAPÍTULO III. EL NUEVO ARTÍCULO 173 BIS CP EN EL PROYECTO DE LEY ORGÁNICA PARA LA PROTECCIÓN DE LAS PERSONAS MENORES DE EDAD EN LOS ENTORNOS DIGITALES 59

1. INTRODUCCIÓN 59

2. BIEN JURÍDICO PROTEGIDO 61

3. SUJETOS DEL DELITO 65

4. CONDUCTA TÍPICA 66

5. ÍTER CRIMINIS 70

6. AUTORÍA Y PARTICIPACIÓN 72

7. PENAS 73

8. PROBLEMAS CONCURSALES 75
9. EL TIPO AGRAVADO DEL PROYECTADO ARTÍCULO 173 BIS 77
CONCLUSIONES 81
BIBLIOGRAFÍA 85
JURISPRUDENCIA 94
NORMATIVA 95

INTRODUCCIÓN

El presente trabajo presenta como objetivo llevar a cabo una aproximación de cómo la difusión o distribución de *deepfakes* sexuales o gravemente vejatorias pueden atentar o poner en peligro bienes jurídicos protegidos, especialmente de las mujeres.

Los avances tecnológicos surgidos en los últimos años han supuesto mejoras significativas en la calidad de vida de las personas, pero también han generado nuevos riesgos y peligros que pueden afectar a derechos fundamentales. En esa línea, la Inteligencia Artificial se ha convertido en una de las herramientas digitales más relevantes y novedosas, con capacidad para incidir en diversos ámbitos de la sociedad.

Entre sus innovaciones más significativas se encuentran las denominadas *deepfakes* o ultrafalsificaciones, contenidos audiovisuales falsos (imágenes, videos o audios) hiperrealistas generados a través de técnicas de *deep learning*. No obstante, dadas sus características pueden emplearse como medios comisivos a efectos de cometer un delito de estafa o de suplantación de identidad, difundir desinformación, alterar los procesos electorales, manipular pruebas en procedimientos judiciales o incluso elaborar y distribuir sin consentimiento contenidos simulados sexuales o vejatorios.

Desde una perspectiva jurídico-penal, los delitos tecnológicos debido a sus características (viralidad, permanencia y anonimato) conllevan un incremento del injusto en comparación con los tipos delictivos que suceden en el mundo analógico. Además, el ilícito penal relativo a la difusión no consentida de *deepfakes* sexuales o vejatorias puede incluso generar un aumento de la lesividad en relación con otros tipos de ataque, debido a la complejidad que supone distinguir entre un contenido real y un contenido falso ya que en ocasiones se trata de materiales manipulados extremadamente sofisticados.

Esta modalidad de delitos afecta desproporcionadamente a las mujeres por varias razones. Por un lado, las despoja del control de su propio cuerpo al tratarlo como un mero objeto susceptible de ser modificado y recreado. Y por otro, pueden emplearse para amenazar o extorsionar, ocasionando así actitudes de control sobre ellas.

El presente estudio tiene como objetivo realizar una primera aproximación de los problemas jurídico penales que puede suscitar el delito de difusión no consentida de *deepfakes* sexuales o gravemente vejatorias -especialmente cuando afectan a mujeres, niñas y niños- elaboradas mediante técnicas de IA generativa.

Para ello el Trabajo de Final de Máster se estructura en tres capítulos. En el primer capítulo expongo una aproximación al término de IA. Para ello, llevo a cabo un análisis de su desarrollo histórico y de sus principales características, centrándome en el tratamiento de las ultrafalsificaciones. A tal efecto, desarrollo el concepto de *deepfakes*, tanto su impacto positivo como negativo en la sociedad, sus modelos de detección y su regulación en el ámbito comunitario y nacional.

El segundo capítulo consiste en desarrollar como las *deepfakes* pueden emplearse como herramienta para ejercer control y dominación sobre las mujeres, convirtiéndose así en una nueva manifestación de violencia de género digital. A tales fines llevo a cabo un breve análisis jurídico penal de otros ciberdelitos, como son el delito de *sexting*, el delito de *grooming* o el delito de *stalking*, entre otros. Además, analizo las características y las diferencias que presentan los supuestos en los que participan simuladamente mujeres políticas o socialmente relevantes o celebridades y aquellos en los que se muestran a mujeres anónimas. Para ello desarrollo qué vinculación puede existir entre la manosfera y la difusión de este tipo de contenidos manipulados, así como con la "pornografía de la venganza" o la sextorsión.

Finalmente, en el tercer y último capítulo me centro en el análisis jurídico penal del proyectado artículo 173 bis CP que recoge el Proyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales. Con vistas a ello examino el bien jurídico protegido y sus problemas de interpretación, los sujetos del delito, la conducta típica, el elemento subjetivo del tipo, el *íter críminis*, las formas de autoría y participación, los posibles problemas concursales que pueden producirse en el futuro, concluyendo con un análisis del tipo agravado del proyectado artículo 173 bis CP párrafo II. Todo ello con el fin de determinar si la propuesta de tipificación es adecuada o no a la necesidad de protección penal en los casos de *deepfakes* pornográficas o vejatorias y detectar así posibles problemas y dudas en torno a su redacción y/o ubicación sistemática.

No obstante, dadas las características y limitaciones del presente trabajo, se ha optado por realizar un análisis genérico de todas y cada una de estas cuestiones, a efectos de establecer una base que permita desarrollarlas con mayor profundidad en futuras investigaciones.

CAPÍTULO I

SOCIEDAD TECNOLÓGICA E INTELIGENCIA ARTIFICIAL

1. INTRODUCCIÓN

Antes de comenzar a analizar los fundamentos de la inteligencia artificial y sus múltiples manifestaciones como medio comisivo de algunos delitos, resulta pertinente explicar y delimitar brevemente el paradigma digital en el que vivimos.

Siguiendo a Dolz Lago (2022:9-11), históricamente, las sociedades han estado en constante evolución. Por ello, desde la aparición del Estado Liberal y en función de los cambios económicos y sociopolíticos, se han ido incorporando paulatinamente múltiples derechos y libertades.

En este contexto, se puede recordar como a cada movimiento revolucionario le ha seguido un reconocimiento de derechos esenciales que suponen avances irrenunciables para los ciudadanos. Como es sabido, en el siglo XVIII, tras la Revolución Francesa, emergieron los derechos humanos de primera generación tales como la libertad de expresión, la de opinión y la de culto, que surgen a efectos de proteger la libertad individual frente a los monarcas absolutos. Posteriormente, en el siglo XIX con la Revolución Industrial, aparecieron los llamados derechos humanos de segunda generación, es decir, los derechos económicos y sociales en respuesta a las precarias condiciones laborales de los trabajadores. Finalmente, en el siglo XX surgió la tercera generación de derechos humanos, también denominados derechos solidarios, entre los que se encuentran, el derecho a la paz, el derecho a disfrutar de un patrimonio común y de un medio ambiente sostenible, con el propósito de garantizar un bienestar colectivo.

En la actualidad, en el siglo XXI y tras el desarrollo de las tecnologías de la información y la comunicación, que dieron origen a lo que ya se conoce como Revolución Tecnológica, se considera que se ha producido una cuarta generación de derechos humanos, los denominados derechos digitales, integrados en lo que Dolz Lago (2022:11) describe como el *"Estado Digital Social y Democrático de Derecho"*. Otros autores, como Serrano Acitores (2023:2), prefieren describir la era digital como la *"cuarta*

revolución industrial". En esa línea existen varios autores que han intentado delimitar qué derechos acompañan a los derechos de la cuarta generación. Precisamente Bustamante Donás (2001:3) sostiene que *"Este conjunto de derechos va tomando forma en las últimas décadas, y abre camino para un gran reto añadido en el siglo XXI: las nuevas formas que cobran los derechos de primera, segunda y tercera generación en el entorno del ciberespacio, es decir, la cuarta generación de los derechos humanos".* En otras palabras, este autor entiende que serían las nuevas manifestaciones de los derechos de la primera, segunda y tercera generación pero en el ámbito digital.

Otros autores, como Villarino Marzo, defienden que los derechos humanos de la cuarta generación se dividen, por un lado, en aquellos derechos relacionados con la biomedicina y la genética (por ejemplo, la integridad genética, la identidad genética, consentimiento informado en todas las intervenciones relacionadas con la salud, etc) y por otro, los derechos que se derivan de las tecnologías de la información y de la comunicación (TIC) (como son el derecho al acceso universal a las nuevas tecnologías o el derecho a la protección de datos, entre otros) (Villarino Marzo, 2017: 55).

En consecuencia, de lo expuesto se constata que la cuarta generación de derechos humanos surge como respuesta frente a los usos ilícitos de las herramientas tecnológicas, que dadas sus características son capaces de generar nuevas formas de atentar contra derechos fundamentales. Todo ello provoca que a través de estos medios se pueda vulnerar o poner en peligro bienes jurídicos protegidos, haciendo necesaria la intervención del Derecho Penal.

Respecto de la IA, como la manifestación más evolucionada de la tecnología podemos hablar de la "singularidad tecnológica", que consiste en un supuesto hipotético en el que el desarrollo tecnológico se vuelve irreversible e imparable, dando lugar a alteraciones o transformaciones impredecibles y de gran entidad para la sociedad (Mucci, 2024).

Junto a ello, la Revolución Digital ha originado una transformación drástica en nuestra sociedad, provocando cambios notables en la comunicación y en las relaciones interpersonales. A través de herramientas y mecanismos como las aplicaciones de mensajería, el correo electrónico y las redes sociales, se establece una comunicación rápida e instantánea con independencia de la ubicación geográfica de los usuarios. Por

ello, la forma en que interactuamos ha evolucionado y resulta sencillo afirmar que, actualmente, la mayoría de las interacciones se realizan en el mundo virtual y, en consecuencia, es imprescindible salvaguardar los derechos de los individuos en el ciberespacio (Matamoros Devís, 2022:1-2). De igual manera, los avances tecnológicos han generado nuevas formas de consumir y de realizar transacciones económicas mediante aplicaciones de venta en línea, como pueden ser Amazon o Shein, entre otras. Las innovaciones tecnológicas impactan también en el ámbito educativo, con el desarrollo de bibliotecas virtuales que conceden el acceso a una gran cantidad de contenidos -libros, revistas, informes, etc- mediante Internet. Igualmente, influyen en la cultura debido a que en muchas ocasiones se emplean para la creación de contenido cinematográfico o para la composición musical, pictórica o literaria.

De este modo, podemos dividir la sociedad en dos *subsociedades*: una analógica y otra digital que tiene lugar en el ciberespacio. En ella, las personas, al igual que en la sociedad física, establecen relaciones personales, económicas y laborales (Lloria García, 2020:63).

Sin embargo, como ocurre con todos los avances, además de producir una gran cantidad de beneficios y progresos, también provocan riesgos y peligros. Tras el desarrollo de la sociedad virtual han surgido nuevos medios comisivos para la perpetración de nuevos ilícitos penales que atentan contra bienes jurídicos tanto individuales como colectivos.

Desde una perspectiva político-criminal existen numerosos problemas que provocan que sea muy compleja la persecución de aquellos delitos que se cometen en el ámbito digital. Entre ellos se encuentra: su facilidad de comisión, el anonimato de los sujetos activos mediante cuentas de usuarios no personales, la consecución de beneficios instantáneos y la transnacionalidad de las acciones al tratarse de una red común (Almenar Pineda, 2018:27).

A su vez, ha adquirido importancia lo que se denomina *"la sociedad de la posverdad",* caracterizada por la difusión de *fake news* o noticias falsas, especialmente en redes sociales a través de cuentas *bots* creadas frecuentemente por sistemas de inteligencia artificial, donde la cultura del *clickbait* o la desinformación prevalecen sobre la verdad objetiva y racional. De igual manera, las innovaciones tecnológicas han

supuesto un riesgo para las mujeres, al emplearse como una herramienta idónea de control y dominación.

En virtud de lo expuesto, resulta fundamental abordar el origen histórico y el concepto de Inteligencia Artificial. En particular, nos centraremos en el estudio de las *deepfakes* o ultrafalsificaciones creadas mediante los sistemas GAN, así como en los riesgos que conllevan y sus métodos de detección. Finalmente, se concluirá con su regulación, tanto en el ámbito comunitario como nacional.

2. IA: DESARROLLO HISTÓRICO Y CONCEPTO

A raíz de los múltiples avances tecnológicos en los últimos años, la Inteligencia Artificial (en adelante IA) se ha convertido en una de las tecnologías más novedosas y relevantes en la actualidad, con la capacidad de incidir en diversas áreas, como la medicina, la ingeniería, la industria cinematográfica e incluso en las ciencias sociales y jurídicas.

El origen de la IA tiene lugar en los años 40, los científicos Warren McCulloch y Walter Pitts han sido reconocidos por ser los pioneros en realizar el primer trabajo relacionado con este término. Para llevar a cabo sus investigaciones se apoyaron en tres aspectos: los conocimientos sobre la fisiología básica y el funcionamiento de las neuronas en el cerebro, el análisis formal de la lógica proposicional de Russell y Whitehead y la teoría de la computación de Turing. Precisamente, Warren McCulloch y Walter Pitts plantearon un modelo formado por neuronas artificiales, en el que cada una de estas se caracterizaba por encontrarse "activada" o "desactivada"; la "activación" tenía lugar como respuesta a la estimulación generada por un conjunto suficiente de neuronas vecinas. Posteriormente, en 1951, dos estudiantes graduados en el departamento de matemáticas de Princeton, llamados Marvin Minksy y Dean Edmons, desarrollaron el primer computador empleando una red neuronal[1] (Russell & Norvig, 2008:19).

[1] Se denominó SNARC (*Stochastic Neural Analog Reinforcement Calculator*), empleaba válvulas de vacío y un mecanismo de piloto automático adquirido de los desechos de un avión bombardero B-24 para simular una red con 40 neuronas (Rusell & Norvig, 2008:19).

Existen diversos trabajos que pueden entenderse que se encuentran relacionados con la IA. Sin embargo, el primer autor en llevar a cabo una visión de la IA fue Alan Turing en 1950 a través de su artículo *Computing Machinery and Intelligence*. En dicho trabajo presentó la prueba de Turing, el aprendizaje automático, los algoritmos genéricos y el aprendizaje por refuerzo (Russell & Norvig, 2008:19-20).

No obstante, no fue hasta el año 1956 cuando el profesor y matemático John McCarthy acuñó el término "Inteligencia Artificial" (*Artificial Intelligence*) en la Conferencia de Dartmouth (*Dartmouth Summer Research Project on Artificial Intelligence*), celebrada en la Universidad de Dartmouth. McCarthy (2007:2) ha definido la Inteligencia Artificial como *"la ciencia y la ingeniería de diseñar máquinas inteligentes, especialmente, programas informáticos inteligentes. Está relacionada con la tarea similar de utilizar ordenadores para comprender la inteligencia humana, pero la IA no tiene por qué limitarse a métodos que sean biológicamente observables".* Entre sus planteamientos, sostenía que la investigación en IA presenta una vertiente teórica y otra vertiente experimental. Asimismo, determinó que existían dos líneas principales de investigación. La primera, de carácter biológico, se inspiraba en la idea de que, dado que los seres humanos son inteligentes, la IA debería estudiarlos e imitar su psicología o fisiología. La segunda, de ámbito fenomenológico, se fundamenta en el estudio y la formalización de hechos de sentido común sobre el mundo, así como los problemas que la sociedad presenta para la consecución o el alcance de objetivos (McCarthy, 2007:12).

En la propuesta del proyecto, McCarthy determinó que la finalidad de la investigación consistía en diseñar un sistema capaz de permitir que las máquinas utilizaran el lenguaje, desarrollaran abstracciones y conceptos, y resolvieran problemas que estaban reservados a la capacidad humana (Nilsson, 2010: 77).

Desde un punto vista gramatical, la Real Academia Española (RAE) define la Inteligencia Artificial como *"la disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o razonamiento lógico"* (Real Academia Española, s.f.). En el mismo sentido, el Alto Grupo de Expertos creado por la Comisión Europea en 2019 elaboró las "Directrices éticas para una IA fiable" en las que describe la IA como una disciplina científica que engloba numerosos enfoques y técnicas, como el aprendizaje

automático (que se compone, por ejemplo, de métodos o técnicas como el aprendizaje profundo y el aprendizaje por refuerzo), el razonamiento automático, que incluye la planificación, programación, representación y razonamiento de conocimientos, así como búsqueda y optimización, y la robótica, que abarca factores como el control, la percepción y el uso de sensores y accionadores, al igual que permite la integración de todas las demás técnicas en sistemas ciberfísicos (Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, 2019).

El término "inteligencia" le otorga diversas capacidades cognitivas que son propias de los seres humanos. De igual manera, el concepto "artificial" se aleja de la subjetividad devenida de sesgos, intereses particulares o planteamientos ideológicos o emocionales, favoreciendo de esta manera que los resultados sean más objetivos y racionales (De La Cuesta Aguado, 2024:87).

Por su parte, el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, conocido como la "Ley de IA", establece que "*la definición debe basarse en las principales características de los sistemas de IA que los distinguen de los sistemas de software o los tipos planteamientos de programación tradicionales o más sencillos (...). Una característica principal de los sistemas de IA es su capacidad de inferencia. Esta capacidad de inferencia se refiere al proceso de obtención de resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos y virtuales, y a la capacidad de los sistemas de IA para deducir modelos o algoritmos, o ambos, a partir de información de entrada o datos".* (Unión Europea, 2024).

Con todo, se puede decir que los sistemas de IA disponen de la capacidad de analizar el entorno que les rodea con una mínima autonomía y adoptar decisiones destinadas a alcanzar objetivos concretos. Estos sistemas pueden ser solamente un programa informático básico (como los sistemas de reconocimiento faciales y de voz, programas de análisis de imágenes o asistentes de voz), o pueden integrarse en dispositivos de *hardware* (por ejemplo, como ocurre con los coches autónomos, robots avanzados o drones) (Comisión Europea, 2018).

Por ello, un sistema de IA se describe como sistema basado en máquinas que pueden llevar a cabo recomendaciones, predicciones o tomar decisiones para un conjunto

determinado de objetivos. Por ello, emplea entradas fundamentadas en máquinas con la finalidad de: 1) percibir entornos reales y/o virtuales; 2) abstraer dichas percepciones previas en modelos de forma manual o automática; y 3) emplear las interpretaciones de modelos para plantear opciones para los resultados (Consejo de Europa, 2019:24).

Por ejemplo, asistentes virtuales como Siri para *Apple* o Alexa para *Amazon* son sistemas que emplean IA a través del reconocimiento de voz de sus usuarios (Hernández Peña, 2022:158). Estos sistemas son capaces de percibir qué es lo que el consumidor desea y, en función de dicha demanda, ejecutar acciones específicas, como puede ser realizar una llamada a una persona en particular, buscar una información concreta en Google o reproducir una canción en una aplicación del móvil o en cualquier dispositivo mediante conexión Bluetooth. Si extrapolamos este proceso al mundo analógico, sería equivalente a cuando una persona le ordena a otra llevar a cabo una acción concreta, como realizar una llamada o reproducir una canción específica en un dispositivo digital. Por lo tanto, estos comportamientos demuestran cómo los programas de IA pueden imitar diversas capacidades humanas.

Existen múltiples clasificaciones en relación con los tipos de IA. Sin embargo, en el presente trabajo nos centraremos en la división clásica. En ella se incluyen la IA fuerte, la IA débil y la IA superinteligente (Pérez-Ugena, 2024:314).

La IA fuerte se puede definir como aquellas máquinas que piensan como los humanos y no únicamente simulan pensar (Ayllón García, 2022:221). Esta clasificación se refiere a los sistemas de inteligencia artificial diseñados para tener la capacidad de simular el pensamiento humano y expresar emociones o conciencia (Hernández Peña, 2022:28). Es decir, dichos sistemas están creados para tener una multitud de habilidades cognitivas y presentar la capacidad de un aprendizaje autónomo (Pérez-Ugena, 2024:314). En consecuencia, tienen la capacidad de aprender, actuar y crear en cualquier situación y, por lo tanto, resolver cualquier tipo de problema, aunque no se encuentre previsto, al igual que lo hace la inteligencia humana (Ayllón García, 2022:221).

La IA débil se define como máquinas que se comportan como si fueran inteligentes (Ayllón García, 2022:221). Consiste en una serie de sistemas diseñados para ejecutar tareas específicas y limitadas, no ostentan capacidad de aprendizaje o adaptación por ellos mismos y necesitan ser programados para llevar a cabo una actividad determinada. Por

consiguiente, no pueden realizar tareas que no se hallen en su ámbito de especialización (Pérez-Ugena, 2024:314). Normalmente, dicha resolución de problemas concretos se realiza empleando técnicas de *machine learning* o *deep learning* (Hernández Peña, 2022:29). Por ejemplo, sería el asistente virtual de Apple Siri (Serrano Acitores, 2023:5).

La IA superinteligente sería una clase de IA que incluso podría llegar a superar la inteligencia humana, al disponer de la habilidad de comprender el mundo y todo aquello que nos rodea a través de capacidades que trascienden las facultades humanas (Pérez-Ugena, 2024:314-315).

Toda esta tecnología permite adelantar la puesta en práctica de acciones que pueden ser lesivas para los derechos fundamentales, especialmente y por lo que se refiere a este trabajo, la IA se ha convertido en un instrumento para elaborar *deepfakes* sexuales o gravemente vejatorias.

3. LAS *DEEPFAKES* O ULTRAFALSIFICACIONES

3.1. Concepto, evolución, usos y funcionamiento

Como he dicho, una de las aplicaciones de la IA es la creación de imágenes realistas que pueden lesionar bienes jurídicos protegidos. Así, es importante fijar el concepto *deepfake* que proviene de los términos en inglés "deep" que equivale "profundo" (y hace alusión al sistema *Deep Learning* o aprendizaje profundo) y "fake" que significa "falso" (García-Ull, 2021:105). En castellano podría traducirse como "ultrafalsificaciones" o "ultrasuplantaciones". Según el diccionario estadounidense Merriam-Webster, el término *deepfakes* se refiere a *"una imagen o grabación que ha sido alterada y manipulada de manera convincente para transgiversar que alguien hizo o dijo algo que en realidad no se hizo ni se dijo"* (Merriam-Webster, s.f.). Este concepto, no obstante, surgió del nombre de un usuario de Reddit[2] que, en 2017, publicó numerosos vídeos pornográficos de

[2] *Reddit* es una red social y un foro en el que sus usuarios se comunican intercambiando ideas, opiniones y pensamientos. La plataforma se estructura en "subreddits" que consisten en foros concretos o "microcomunidades" en las que sus usuarios debaten, difunden o comparten contenidos audiovisuales de una materia en particular. Asimismo, los usuarios votan positivamente o negativamente tanto los comentarios como las publicaciones compartidas y de esta manera la plataforma clasifica los contenidos en relación con los votos obtenidos. Disponible en: https://www.reddit.com/.

celebridades como Taylor Swift o Scarlett Johansson (Simó, 2023:496 y Kweilin, 2022: 650).

Desde un plano normativo, el Reglamento (UE) 2024/1689 del Parlamento Europeo y el Consejo, define las *deepfakes* como *"un sistema de IA para generar o manipular un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeje notablemente a personas, objetos, lugares, entidades o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos"*. En otras palabras, los *deepfakes* se crean mediante el uso de tecnologías de procesamiento de gráficos y técnicas avanzadas de aprendizaje profundo, como son las redes neuronales recurrentes (RNN[3]) o redes generativas adversarias (*generative adversarial networks* o GAN[4]) (Barba Arteaga, 2024:231), con el propósito de generar contenidos visuales hiperrealistas en los que se representen de manera simulada a personas que han dicho o han llevado a cabo acciones o manifestaciones que realmente nunca sucedieron ni dijeron (García-Ull, 2021:107). Durante este procedimiento, el rostro de una persona es analizado por un algoritmo de aprendizaje profundo (*deep learning*) y luego se superpone sobre el rostro de otra persona; esto se logra dado que el algoritmo ha "aprendido" la fisonomía del rostro, sus movimientos y las expresiones de la persona (Öhman, 2019:133).

A su vez, uno de los presupuestos más relevantes de la tecnología *deepfake* es el aprendizaje automático (*machine learning*). En dicho procedimiento, se utilizan un conjunto de redes artificiales (RNA) denominadas redes convolucionales (CNN) (Okolie, 2023:5). Las RNA están diseñadas para imitar la inteligencia humana, ya que se fundamentan en el funcionamiento de los sistemas nerviosos biológicos, como es el cerebro humano (O´Shea y Nash, 2015:1-2).

En particular, los *deepfakes* se producen mediante los sistemas de redes generativas antagónicas/adversarias o GAN (*generative adversial network*), un tipo de algoritmo de aprendizaje que se caracterizan por entrenar dos redes neuronales (una red generadora y una discriminadora) que se adiestran conjuntamente para competir simultáneamente de

[3] Una red neuronal recurrente (RNN) son una clase de aprendizaje profundo que se fundamenta en los trabajos de David Rumelhart en 1986. Las RNN se caracterizan por ostentar capacidad para procesar y obtener información de datos secuenciales. Por ese motivo, dependen de redes neuronales recurrentes acciones como el análisis de vídeo, la subtitulación imágenes, el procesamiento del lenguaje natural (PLN), así como el análisis de la música (Arana, 2021:5).

forma antagónica. La red generadora o generador se encarga de producir la imagen creada por la IA, al tiempo que la red discriminadora o discriminador, mediante un proceso de entrenamiento, trata de diferenciar entre las imágenes reales y los datos creados por el generador (Vestri, et al., 2023:287-288).

Con esta técnica, se producen diferentes clases de manipulaciones que Santiesteban (2024:216) fundamentándose en el informe *Snapshot Paper Deepfakes and Audiovisual Disinformation* elaborado por el *UK Centre For Data Ethics And Innovation*, clasifica así:

1) El *face replacement* que consiste en tomar una imagen del rostro de una persona específica y superponerla sobre el rostro de otra.

2) El *face reenactment* implica recrear rostros modificando diferentes características pero no conlleva un cambio de cara.

3) El *face generation* hace referencia a la generación de rostros, que simulan caras, pero no se corresponden con ningún rostro real o existente.

4) El *speech synthesis* se trata de aquellos procesos por los que se crean modelos de voz mediante el empleo de un sistema de *software* avanzado a efectos de generar un modelo de voz de una persona específica.

La tecnología *deepfake* se emplea de manera significativa en la industria cinematográfica, especialmente en la creación de clones y en la modificación y creación de personajes en películas (Okolie, 2023:6). Asimismo, este tipo de tecnología de IA también ha incidido en el ámbito de la publicidad, a efectos de captar la atención de potenciales clientes y consumidores. Un ejemplo de ello es la utilización de la recreación de Lola Flores en una campaña publicitaria de una marca de cervezas española[5].

Por otro lado, también se están incorporando en los museos de arte. Por ejemplo, el Museo Dalí de San Petersburgo (Florida, EE.UU) organizó una exposición denominada *Dali Lives,* en la que, a través de enormes pantallas una imagen animada de Dalí interactuaba con el público (Mihailova, 2021:884). Debido a las características de los contenidos, estos frecuentemente son subidos o difundidos a las redes sociales como

[5] Anuncio disponible en: https://www.google.com/search?q=anuncio+de+lola+flores+curzcampo&rlz=1C5CHFA_enES898ES899&oq=anuncio+de+lola+flores+curzcampo&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIJCAEQABgNGIAEMgoIAhAAGIAEGKIE0gEINzc3OWowajeoAgCwAgA&sourceid=chrome&ie=UTF-8#fpstate=ive&vld=cid:2820cf72,vid:Yewm6TfLZ3Q,st:0. (Recuperado el 13 de mayo de 2025).

Instagram, X o TikTok, donde últimamente se están viralizando contenidos *deepfakes* satíricos de figuras políticas relevantes como Donald Trump o Barack Obama[6], vídeos que muestran cómo serían determinados personajes históricos en la actualidad[7], entre otros.

No obstante, dado que se trata de un medio tecnológico, la licitud de su utilización dependerá del propósito de quién lo emplee. Mediante el uso de *deepfakes,* se puede manipular la voz o elaborar vídeos con el fin de cometer un delito de estafa, también se pueden generar vídeos de políticos de carácter populista con la finalidad de alterar la opinión pública en procesos electorales e, incluso, es posible la creación de contenido pornográfico sin consentimiento a través de imágenes manipuladas. Es decir, diversas situaciones que nos lleva al análisis de los riesgos que pueden generar este tipo de técnicas de IA generativa.

3.2. Riesgos generales derivados de las *deepfakes*

Como acabo de exponer, las *deepfakes*, o cualquier otro producto de la tecnología, tienen aspectos positivos y negativos. En ese sentido, la creación de contenidos simulados puede afectar a distintos bienes jurídicos según el tipo de acción que se produzca. Concretamente, pueden menoscabar el Estado de derecho al amenazar los procesos democráticos mediante la posibilidad de manipular la opinión pública, también tienen la capacidad de intervenir en el acceso a la justicia puesto que se pueden aportar pruebas falsas generadas mediante IA generativa, además, constituye un instrumento idóneo para cometer diversos delitos económicos. Finalmente, pueden ser un peligro para la integridad moral, principalmente de las mujeres, mediante la creación de contenidos pornográficos o vejatorios, actuando como un nuevo medio de control y dominación, y todo ello simplemente a título de ejemplo.

[6] Contenidos disponibles en: https://www.youtube.com/watch?v=cQ54GDm1eL0 y https://vm.tiktok.com/ZNdddaTUu/. (Recuperado el 13 de mayo de 2025).
[7] Contenido disponible en: https://vm.tiktok.com/ZNdddXdYg/. (Recuperado el 13 de mayo de 2025).

3.2.1. Impacto de las deepfakes en el ámbito político y en la difusión de desinformación

Actualmente, vivimos en una época caracterizada por una transformación permanente y constante, fundamentada esencialmente por la información y los avances tecnológicos y digitales. Este progreso ha impulsado la creación de la denominada *Sociedad de la Información*. Este término alude a la importancia social que se le otorga a la información y la comunicación en la sociedad actual, donde se incluyen las relaciones culturales, sociales y económicas (Alfonso Sánchez, 2016:236).

El avance tecnológico ha provocado que la IA se expanda e incluso incida en ámbitos tan relevantes de la sociedad como son las campañas políticas o los procesos electorales. Todo ello, ha originado que surjan tanto ventajas como riesgos. Entre los usos beneficiosos se encuentran la creación de *chatbots* con capacidad de proporcionar información en tiempo real sobre los colegios electorales, los procedimientos de votación o las plataformas de candidatos. También la IA puede optimizar la gestión de datos al llevar a cabo la recopilación, almacenamiento, así como un análisis minucioso de un gran número de datos electorales (Naciones Unidas, 2024). Estas actuaciones han impulsado que el proceso electoral sea más transparente y más accesible para los ciudadanos y, en consecuencia, se obtenga una mayor precisión de los votos, otorgando así una mayor calidad democrática.

Sin embargo, en este contexto la desinformación juega un importante papel. Esta podría definirse como *"la difusión de información deliberadamente falsa"* (Oxford English Dictionary, s.f.). En otras palabras, la desinformación es un fenómeno en el que el emisor tiene la intención de manipular o influenciar a los receptores con el objetivo de que actúen según sus intereses (Rodríguez Andrés, 2018:235)

Por lo tanto, entre los riesgos que se pueden producir, se encuentra la posibilidad de elaborar *fake news* mediante la tecnología *deepfake*, en las que parezca que una persona ha dicho o hecho algo que realmente nunca sucedió. Estas actuaciones tienen como objetivo desacreditar a los respectivos candidatos políticos o difundir noticias falsas que influyan directamente en la opinión pública, generando, bien un engaño en los receptores al creer que esas imágenes son reales o bien incertidumbre al no poder diferenciar si ese contenido es verdadero o falso (Vaccari & Chadwick, 2020:3-4). En el

caso analizado estas acciones también pueden acabar con la reputación de la persona en otros ámbitos. Sin embargo, ¿de qué manera podría afectar un *deepfake* a los procesos electorales y, en consecuencia, amenazar la democracia?

Por ejemplo, imaginemos que, a pocos días del comienzo de un proceso electoral, se difunde por una red social un vídeo falso de un candidato político en el que realiza o comunica algo inapropiado que provoca una desacreditación hacia él y hacia su ideología. Si dicho contenido manipulado se viraliza podría repercutir en la opinión de los electores y, en consecuencia, influir en su éxito electoral. Concretamente, estos contenidos manipulados pueden perjudicar su reputación, así como indirectamente puede ocasionar una desconfianza general en los políticos involucrados y una confusión sobre si una información o noticia es real o falsa (Chesney & Citron, 2018:1777, Vaccari & Chadwick, 2020:2, Aguayo & Ruiz del Olmo, 2024:5).

Una situación parecida ocurrió en otoño de 2023, en vísperas de las elecciones en Eslovaquia. En ese sentido, se difundieron por redes sociales unas grabaciones de audio falsas en las que se escuchaba al líder del partido prooccidental y uno de los más importantes del país, "Eslovaquía Progresista", discutir la forma de manipular la votación y proponer un incremento del precio de la cerveza. En esas elecciones "Eslovaquia Progresista" fue derrotada, algunos expertos señalan que puede haber sido las primeras elecciones de la historia manipuladas a través de contenidos ultrafalsos (De Tena Piera, 2024). Otro caso relevante ocurrió en la India. Previamente a las alecciones se difundieron y se viralizaron dos vídeos que mostraban a dos estrellas de Bollywood reconocidas, Ranveer Singh y Amir Khan, promover el apoyo a la campaña política del partido opositor. Ambos actores denunciaron que se trataba de un vídeo *deepfake* realizado sin su consentimiento (BBC, 2024).

En esa línea, la desinformación también ha desempeñado un papel relevante durante la campaña de Donald Trump en las elecciones estadounidenses de 2016, también durante el referéndum del Brexit en Reino Unido, así como en las elecciones europeas de 2019 en Italia (Gil Lloria, 2025:86-87). Por lo tanto, de todo ello se deduce que las *deepfakes* generadas mediante IA (ya sean imágenes, audios o vídeos hiperrealistas pero falsos) pueden emplearse con el objetivo de engañar a los votantes, socavando el proceso electoral y la democracia. En ese sentido, García-Ull (2021:110) expone que incluso los

deepfakes representan mayor riesgo que las noticias falsas tradicionales ya que resultan más complejos de detectar.

Los riesgos de la desinformación se incrementan en la sociedad digital ya que las redes sociales tienen la capacidad de poder difundirlos entre una audiencia amplia de personas de forma rápida, inmediata y generalizada. Precisamente, la difusión conlleva grandes riesgos puesto que depende de varios factores. Por un lado, de las personas que a consciencia generan ese contenido falso y lo publican y, por otro, aquellas personas que, han sido engañadas y creyendo que el contenido es verdad, lo difunden (Farenc Lendvai & Gosztonyi, 2024:4). Esta situación plantea el problema sobre cuáles son aquellos elementos que pueden provocar una confusión en los receptores.

El engaño puede producirse por varias razones: bien por la capacidad de las GAN de generar vídeos hiperrealistas en los que resulta complicado detectar lo verdadero de lo falso, o porque el contenido manipulado puede haber conseguido una gran cantidad de *"me gustas"* y reproducciones tanto de usuarios reales como de cuentas *bots* creadas por IA que otorguen esa apariencia de veracidad, o porque resulta probable que algunos dirigentes políticos o creadores de contenido con miles de seguidores y con incidencia en la sociedad, compartan esa información manipulada. En ese sentido, Vaccari y Chadwick desarrollaron una investigación en la que evaluaron la capacidad de las *deepfakes* para inducir a sus receptores al engaño. Concretamente, un vídeo de cuatro segundos engañó al 15% de los participantes y un 35% no estaban seguros de su autenticidad, un vídeo de 26 segundos engañó a un 16% y un 37% se mostraban inseguros (Vaccari & Chadwick, 2020, 5-8 y Farenc Lendvai & Gosztonyi, 2024:4).

Otro elemento que conlleva un riesgo y provoca que exista un gran número de contenidos manipulados en redes sociales es la facilidad de su creación. Actualmente, no es necesario poseer grandes conocimientos técnicos ni informáticos, ya que existen varias aplicaciones como, *FakeApp, Faceswap* o *DeepFace Lab* que permite a cualquier usuario elaborar *deepfakes* de manera inmediata y rápida (Barba Arteaga, 2024: 236-237 y Paris & Donovan, 2019:14). Aunque, en ocasiones, estas aplicaciones no generan materiales muy sofisticados, pueden ser suficientes para generar incertidumbre sobre su veracidad.

Todas estas actuaciones pueden originar una distorsión de la democracia, ya que tienen la capacidad de incidir en los procesos electorales[8], disminuir la confianza en las instituciones, provocar un aumento de las divisiones sociales (Citron & Chesney, 2018:1777-1780), así como, una pérdida de confianza en las noticias que compartidas en las redes sociales (Vaccari & Chadwick, 2020:3), especialmente entre los jóvenes.

3.2.2. Las *deepfakes* como instrumento para cometer delitos económicos

Entre los riesgos más significativos que originan las ultrafalsificaciones encontramos aquellos que se destinan a cometer delitos de carácter económico. Todo ello es debido a que las *deepfakes* permiten clonar la voz de individuos o crear contenidos audiovisuales en los que muestran a personas que nunca han dicho o hecho lo que se reproduce. Por lo tanto, esta tecnología se ha convertido en una herramienta idónea para perpetrar delitos de estafas muy sofisticadas y con una capacidad de engaño considerable.

El Foro Económico Mundial ha constatado que en 2022, el 66% de los profesionales de ciberseguridad sufrieron ataques *deepfakes* dentro de sus respectivas organizaciones (Foro Económico Mundial: 2022).

El fraude se fundamenta en la capacidad de las *deepfakes* de recrear o superponer rostros, clonar voces y gestos de otra persona, generalmente de figuras de confianza o personas conocidas para la víctima, como el CEO de su empresa, algún familiar o un compañero de trabajo, entre otros (SEON, glosario:2025). Por lo tanto, existen varias formas en las que los estafadores podrían emplear la tecnología *deepfakes* a efectos de cometer cualquier delito de carácter económico. Estos pueden perpetrar, por ejemplo, un fraude fantasma*,* es decir, mediante un *deepfake* un individuo se hace pasar por alguien que ha fallecido para conseguir un traspaso patrimonial o el fraude de cuenta nueva si un sujeto, a través de contenidos manipulados, suplanta la identidad de otra persona para abrir cuentas bancarias. Y finalmente, el fraude de identidad sintética en el que un sujeto emplea las ultrafalsificaciones para crear una persona que en realidad no existe y así poder

[8] De igual manera, el uso de ultrafalsificaciones puede incidir en los conflictos armados. En el año 2022 unos piratas informáticos con el propósito de crear miedo y confusión provocaron que apareciera en el canal de noticias "Ucrania 24", un vídeo manipulado del presidente ucraniano Volodímir Zelenski comunicando a los soldados que depusiera las armas y se rindieran (Allyn, 2022).

solicitar tarjetas de crédito o débito u otras transacciones que nunca devolverá (Álvarez, 2024).

En la actualidad existen ya casos de estafas cometidas mediante el uso de *deepfakes*. Un ejemplo significativo sucedió en 2023, cuando unos estafadores de Arizona (EE.UU.) intentaron engañar a una mujer a través de una llamada telefónica. La mujer escuchó la voz de su hija de 15 años solicitando ayuda y mientras los estafadores le pedían un rescate económico. Por fortuna, la madre logró descubrir que su hija se encontraba bien, averiguando así que estaba siendo víctima de una estafa (Anikin, 2023).

Otro ejemplo de fraude ocurrió en una empresa multinacional. Un trabajador fue convocado a una reunión en línea. Al conectarse contempló que estaban presentes el director de finanzas y otros superiores. Sin embargo, no eran ellos en realidad sino recreaciones sintéticas. Los ciberestafadores habían suplantado la identidad de estos sujetos a efectos de ganarse su confianza. El trabajador engañado ejecutó las órdenes que les proferían, entre ellas, la realización de una transferencia bancaria de unos 200 millones de dólares a una cuenta no vinculada a la empresa (La Vanguardia, 2024 y Kong, 2024).

Por otro lado, la herramienta tecnológica de las *deepfake* también puede ser empleada para favorecer la vulneración de las interfaces de programación de aplicaciones de reconocimiento facial y, dadas sus funciones, se puede desbloquear medidas de seguridad necesarias para acceder al teléfono móvil o realizar una transferencia bancaria, como por ejemplo ocurre con el sistema de pago de Apple que necesita de verificación facial para poder llevar a cabo un pago (Blázquez, 2023:231). Considerando que mediante la tecnología *deepfakes* es posible recrear rostros, expresiones y características faciales se puede vulnerar las medidas de seguridad propias de las aplicaciones de entidades bancarias o del teléfono móvil y cometer así un delito de carácter económico.

En ese sentido, EUROPOL (2020:60) también advierte que, además de cometer estafas económicas, las *deepfakes* pueden facilitar la revelación de secretos o datos confidenciales de empresas o de particulares. Por ejemplo, imaginemos que, a través de una videollamada o una llamada ordinaria, un individuo se hace pasar por un dirigente o por un compañero de la empresa, instando a que le proporcionen cualquier tipo de información de carácter confidencial. Por consiguiente, como la víctima confía en dicho sujeto puede llegar a comunicárselo, revelando así un dato que en realidad es secreto. Es

decir, de lo expuesto se puede deducir que las posibilidades de lesión de bienes jurídicos protegidos de naturaleza económica son muchas y diferentes.

3.2.3. Utilización de las *deepfakes* en la manipulación de pruebas en procesos judiciales

A lo largo del presente estudio se ha analizado que el desarrollo de la era digital, y en particular la aparición de nuevas herramientas como las ultrafalsificaciones, presentan la capacidad de incidir en numerosos ámbitos, entre los cuales se encuentran la presentación de pruebas manipuladas en los procedimientos judiciales.

Antes del análisis de dicha cuestión, es necesario diferenciar, por un lado, entre aquellos contenidos audiovisuales en los que los sistemas de IA han mejorado su calidad y, en consecuencia, no conllevarían ningún inconveniente procesal ni penal. Y por otro, aquellos que han sido manipulados con el propósito de engañar al juez en un proceso judicial. Todo ello plantea la siguiente cuestión: ¿de qué manera un contenido *deepfake* puede introducirse en un procedimiento judicial?

Desde una perspectiva procesal, en virtud del artículo 299 de la Ley de Enjuiciamiento Civil y el artículo 26 del Código Penal, las ultrafalsificaciones en sus diversos soportes (imagen, vídeo y audio) podrían presentarse como medio de prueba documental en formato digital (Blázquez, 2023:235-236). La aportación y detección de una prueba manipulada creada mediante sistemas GAN puede desencadenar una serie de consecuencias jurídico-penales, como un delito de estafa procesal (artículo 248 del Código Penal) o un delito de falsedad documental (artículo 395 del Código Penal).

La tecnología *deepfake* ostenta la capacidad de clonar voces en archivos de audio, superponer rostros y falsificar gestos. Por lo tanto, es una herramienta idónea para introducir pruebas manipuladas en procedimientos judiciales y, de este modo, vulnerar el derecho fundamental a la tutela judicial efectiva recogido en el artículo 24 de la Constitución Española. En particular, un *deepfake* podría ser aportado en el proceso junto con el dispositivo de almacenamiento en el que se halle y, por lo tanto, la prueba se llevará a cabo sobre el *hardware*. También sería posible introducirla en formato intangible, es decir, el producto que resulta del sistema de elaboración de ultrafalsificaciones. En tal

caso, se le remitirá el archivo al LAJ y, posteriormente, su contenido será transcrito en un soporte duradero a efectos de facilitar su manejo (Blázquez, 2023:238).

Según la UNESCO (2023:35), en su informe denominado *"Kit de herramientas global sobre IA y el estado de derecho para el poder judicial",* dispone que la incorporación de pruebas generadas IA plantea varios desafíos, riesgos e implicaciones jurídicas, tales como la admisibilidad por parte de los tribunales de las pruebas elaboradas por la IA, el testimonio de expertos. De hecho, el tribunal puede enfrentarse al paradigma de si la IA puede entenderse como un "testigo experto" y como se ha de tratar su respectivo testimonio, también se puede ocasionar el surgimiento diversas consecuencias éticas y la necesidad de una supervisión humana a efectos de analizar la autenticidad de las pruebas y la imparcialidad de las respectivas decisiones legales.

En ese sentido, resulta fundamental que el poder judicial comience a invertir en medios que faciliten la detección de contenidos generados a través de IA. Por ejemplo, sería interesante la elaboración de un informe pericial que demuestre la veracidad de la prueba aportada (Montesinos García, 2024:580-581) y así evitar resoluciones injustas basadas en pruebas falsas.

3.2.4. Empleo de las *deepfakes* para la creación de pornografía. Especial referencia a la pornografía infantil (remisión)

Entre los usos más comunes de las *deepfakes* se hallan aquellas que se elaboran con el propósito de crear contenido pornográfico y/o vejatorio. En ese sentido, podrían darse tres escenarios:

i) cuando el contenido pornográfico se crea con el consentimiento de la/s persona/s que aparecen en la imagen, por lo que en tal caso no constituye un ilícito penal

ii) cuando el contenido se elabora sin el consentimiento de la/s personas implicada/s y, por lo tanto, se produce una actuación reprochable penalmente

y iii) cuando el contenido muestra una persona menor de edad, lo que podría constituir un delito de pornografía infantil virtual tipificada en el artículo 189 CP.

Los *deepfakes* pornográficos se elaboran a través de los sistemas de IA generativa que permiten superponer el rostro de una persona, generalmente una mujer, en un cuerpo

de una actriz porno. Mediante las ultrafalsificaciones se muestra a una persona que parezca que está realizando un acto de carácter sexual que realmente nunca ha llevado a cabo. Concretamente, según un informe desarrollado por *Deeptrace* (2019:1-2), de 14.678 vídeos manipulados en línea, el 96% era contenido pornográfico y en el 100% de ellos participaban mujeres. Otro informe, denominado 2023 *State of deepfakes* llevado a cabo por *Home Security Heroes* se determinó que la pornografía *deepfake* constituye un 98% de todos los vídeos *deepfakes* en línea y el 99% de las *deepfakes* pornográficas se encuentra protagonizado por mujeres. Por lo tanto, este tipo de materiales se utilizan para controlar a las mujeres y limitar su representación en la sociedad (Jacobsen y Simpson, 2024:1100).

No obstante, existen dos perspectivas en función del sujeto que participa en esos contenidos simulados. En primer lugar, generalmente si la víctima es una mujer anónima, el hecho se encuentra ligado al término "pornografía de la venganza" (Barba Arteaga, 2024:234). En tal caso, se utiliza dicha herramienta tecnológica como forma de chantaje emocional ya sea para la consecución de favores sexuales o monetarios (Okolie,2023:4). Imaginemos que una mujer deja a su novio o esposo y él la chantajea con que, si no vuelven a ser pareja, difundirá un vídeo *deepfake* en el que aparece ella realizando un acto sexual.

En segundo lugar, cuando la víctima del vídeo *deepfake* es una celebridad o una mujer con relevancia social y/o política, suele obedecer a diferentes patrones. En este supuesto ya no se aprecia esa venganza o ese chantaje por parte del varón hacia la mujer, sino un ánimo de humillación o gratificación sexual para los hombres. Estas actuaciones se hallan estrechamente relacionadas con la manosfera o machosfera, es decir, comunidades de hombres que mediante aplicaciones para móviles o redes sociales difunden contenidos, discursos y acciones sexistas o misóginas (Lloria García, 2024:77-78). Por esta razón, en este tipo de comunidad es factible que se elaboren contenidos *deepfakes*, ya sea de celebridades o de mujeres relevantes, con el objetivo de humillarlas u obtener placer sexual, atentando así contra su integridad moral, como sucedió con Taylow Swift o Scartlett Johansson (Simó Soler, 2023:496 y Kweilin, 2022:650), entre otras. En cambio, cuando la víctima es una mujer con relevancia social y/o política, suele ir acompañado de un ánimo de menospreciarlas y humillarlas, como ocurrió con Rana

Ayyub, Kate Isaacs o Cara Hunter. En ese sentido, en ambos hechos se estaría atentando contra la dignidad de las mujeres por el mero hecho de serlo.

Por todo ello, resulta conveniente señalar que no únicamente las *deepfakes* pornográficas o sexuales pueden atentar contra la integridad moral de las mujeres. Por ejemplo, un *deepfake* en el que las mujeres sean tratadas de forma gravemente vejatoria, aún sin practicar una conducta sexual, podrá atentar contra dicho bien jurídico.

En definitiva, las ultrafalsificaciones se han convertido en una nueva forma de control y dominación sobre las mujeres, reduciéndolas a meros objetos sexualizados y consumibles. Y de esta manera, aparece una nueva manifestación de la violencia de género, dirigida a perpetuar la discriminación de las mujeres, en este caso en el ámbito digital, tal y como veremos más adelante[9].

4. MÉTODOS DE DETECCIÓN DE LAS *DEEPFAKES*

El aumento de la creación de *deepfakes* y su difusión mediante las Tecnologías de la Información y la Comunicación (TIC) ha generado el desarrollo y la progresiva actualización de diversas herramientas destinadas a la detección de los contenidos manipulados. Estas herramientas pueden estar basadas en IA, también existen herramientas basadas en el análisis de metadatos, en la fecha de creación del archivo o la ubicación de la grabación pueden disponer de pistas que hagan sospechar que se trata de un contenido manipulado y finalmente, existen también las herramientas basadas en la detección de anomalías, este tipo de herramientas presentan como objetivo la búsqueda de patrones inusuales en la voz, el movimiento facial o del rostro o el contenido en si del video que apunte a que se trata de una ultrafalsificación (Micucci, 2024).

La detección de contenidos manipulados por las GAN puede llevarse a cabo mediante la percepción humana o a través del uso de diversas herramientas tecnológicas destinadas exclusivamente a esta cuestión. De este modo, las capacidades humanas podrían detectar, por ejemplo, las *cheapfakes*, es decir, contenido manipulado mediante

[9] Esta cuestión se analiza con detenimiento en el Capítulo II.

IA pero menos sofisticadas que las *deepfakes* puesto que suelen presentar errores visibles y detectables para cualquier persona.

No obstante, existen contenidos manipulados hiperrealistas en los que se necesita una herramienta especializada para determinar si se trata de un contenido real o falso. Para ello se han elaborado numerosos métodos con el propósito de detectar la autenticidad de los materiales. Generalmente se suelen emplear herramientas de IA ya que las herramientas de creación de *deepfakes* necesitan datos de entrenamiento para saber cómo es una persona real, la mayoría de los modelos de detección de *deepfake* se entrenan utilizando bases de datos de ultrafalsificaciones (Europol, 2024:17).

Estas herramientas que emplean IA utilizan algoritmos de aprendizaje automático destinados a analizar patrones en audios o vídeos con la finalidad de identificar y buscar inconsistencias que indiquen que ese contenido ha sido manipulado (Micucci, 2024). Por ejemplo, herramientas como *Tensorflow* y *PyTorch* utilizan redes neuronales profundas a efectos de detectar falsificaciones en vídeos, imágenes y audios (Salako Emmanuel, 2023). De igual manera, la biometría ha resultado ser una herramienta muy eficaz para la detección de *deepfakes*, ya que, mediante sofisticados algoritmos y técnicas de aprendizaje automático, el sistema tiene la capacidad de analizar características de la persona, tales como su estructura facial, el iris de los ojos o los patrones de voz. Todo ello, con el propósito de buscar posibles errores o incongruencias que evidencien que el contenido es simulado (Nebrada, 2023).

La entidad de *software* y servicios biométricos, *Aware Biometrics,* expone que la tecnología de autenticación biométrica destinada a la detección de contenidos falsos ofrece diversos servicios o métodos como: i) reconocimiento facial, basado en el análisis de las características del rostro de un individuo (tamaño de la nariz, color del iris de los ojos, rasgos de la boca, entre otras); ii) reconocimiento de voz, que evalúa los distintos aspectos que tiene la voz de una persona, como el ritmo o el tono, con el propósito de identificar patrones de habla que no sean reales; iii) la biometría del comportamiento analiza los patrones corporales que emplea un individuo en concreto; y iv) la biometría multimodal, en la que se combinan varios métodos de autenticación biométrica a efectos de aumentar la seguridad (Aware Biometrics, 2024).

5. REGULACIÓN DE LAS *DEEPFAKES* EN EL ÁMBITO COMUNITARIO Y NACIONAL

El avance y el desarrollo de la IA ha provocado que la UE y sus estados miembros tomen conciencia de la trascendencia de la situación y comiencen a elaborar progresivamente textos legislativos con el propósito de regular las actuaciones analizadas.

Para ello, en los últimos años, la Unión Europea ha ido estableciendo un marco normativo a efectos de controlar el creciente desarrollo de las nuevas tecnologías, y en particular de la Inteligencia Artificial. En el año 2018 la Comisión Europea emitió la COM (2018) 237 o "*Artificial Intelligence for Europe*", de 25 de abril y la COM (2018) 795, de 7 de diciembre o "Plan coordinado sobre inteligencia artificial". Estas Comunicaciones presentaban como finalidad elaborar las bases legislativas para una posterior regulación coordinada entre Estados Miembros sobre la IA. En ese sentido, en abril de 2019, la Comisión Europea publicó la COM (2019) 168, de 8 de abril o "Comunicación: generando confianza en la inteligencia artificial centrada en el ser humano".

Asimismo, la Comisión creó un Grupo de Expertos de Alto Nivel sobre la IA a efectos de asesorar sobre esta materia. Por ello, se elaboraron tres documentos relevantes: "*Las Directrices éticas para una IA fiable"*, "*Recomendaciones políticas y de inversión para una IA fiable*" y "*Lista de evaluación para una IA fiable".*

Posteriormente, en febrero del año 2020 mediante la COM (2020) 65 final, se desarrolló "el Libro Blanco sobre Inteligencia Artificial", cuya finalidad consiste en sentar las bases sobre las medidas que adoptará la UE en relación con la IA. Más tarde, la Comisión emitió la COM (2021) 205 final, de 21 de abril o "Comunicación sobre el fomento de un enfoque europeo de la IA", en la que se presentó un marco normativo. Finalmente, en marzo de 2024, se adoptó el texto legislativo más relevante: el Reglamento (UE) 2024/1689, del Parlamento Europeo y del Consejo, de 13 de junio de 2024 también denominado "*Reglamento IA*" o "*Ley IA*".

El Reglamento (UE) 2024/1689, recoge en su artículo primero[10] que su objetivo consiste en mejorar el funcionamiento del mercado interior mediante el establecimiento

[10] El artículo primero del Reglamento (UE) 2024/1689 señala:

un marco jurídico uniforme destinado al desarrollo, la introducción en el mercado, la puesta en servicio y la utilización de sistemas de inteligencia artificial a efectos de adoptar una IA centrada en el ser humano y fiable, todo ello tomando como eje y fundamento los valores de la Unión. La Ley IA desarrolla una clasificación de los sistemas de IA según su nivel de riesgo. Es decir, en función de cómo se aplique y cómo se utilice dicha herramienta puede generar riesgos y menoscabar los intereses públicos y los derechos fundamentales de la UE. Esta afectación puede ser tangible o intangible, pudiendo provocar daños físicos, psíquicos, económicos o sociales (Gil Cruz, 2024:1). El Reglamento divide el riesgo en cuatro tipos: riesgo inaceptable, riesgo alto, riesgo bajo o riesgo mínimo.

Los sistemas catalogados como bajo o limitados están sujetos a unas normas de transparencia; los sistemas de riesgo alto han de cumplir unas obligaciones y unos presupuestos rigurosos para poder acceder al mercado de la UE y finalmente, los designados como riesgo inaceptable se hallan prohibidos por la UE (Consejo de Europa y Consejo Europeo, 2025). En el Reglamento (UE) 2024/1689 se menciona explícitamente las *deepfakes*. En él se desarrolla una definición del concepto[11] y, por ende, debido a su ubicación sistemática se entiende que lo identifica como hechos de riesgo bajo o limitado. Por esta razón, se regula en el artículo 50.4[12] recogido en el Capítulo IV sobre "Obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA". Este precepto establece que todos aquellos

> *"El objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior y promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA (en lo sucesivo, «sistemas de IA») en la Unión así como prestar apoyo a la innovación".*

[11] En el apartado sesenta del artículo 3 del Reglamento (UE) 2024/1689, se definen a las ultrafalsificaciones o *deepfakes* como *"un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeja a personas, objeto, lugares, entidades o sucesos reales y que pueden inducir a una persona a pensar erróneamente que son auténticos o verídicos".*

[12] El artículo 50.4 del Reglamento (UE) 2024/1689 señala que:

> *"Los responsables del despliegue de un sistema de IA que genere o manipule imágenes o contenidos de audio o vídeo que constituyan una ultrasuplantación harán público que estos contenidos o imágenes han sido generados o manipulados de manera artificial. Esta obligación no se aplicará cuando la ley autorice su uso para para detectar, prevenir, investigar o enjuiciar delitos. Cuando el contenido forme parte de una obra o programa manifiestamente creativos, satíricos, artísticos, de ficción o análogos, las obligaciones de transparencia establecidas en el presente apartado se limitarán a la obligación de hacer pública la existencia de dicho contenido generado o manipulado artificialmente de una manera adecuada que no dificulte la exhibición o el disfrute de la obra".*

autores de contenidos de imagen, audio o vídeo, generados o manipulados por IA han de indicar que han sido creados mediante este tipo de sistemas salvo cuando la ley autorice su uso para detectar, prevenir, investigar o enjuiciar delitos.

Sin embargo. el Reglamento no parece entender que las *deepfakes* supongan un riesgo alto a diferencia de lo que recoge la Directiva (UE) 2024/1385, de Parlamento Europeo y del Consejo, de 14 de mayo de 2024, sobre la lucha contra la violencia contra las mujeres y la violencia doméstica, que insta a los Estados miembros a tipificar estas conductas como ilícito penal. De tal modo, este texto normativo presenta como propósito proporcionar un marco integral a efectos de prevenir y combatir de forma eficaz la violencia contra las mujeres y la violencia doméstica en la Unión Europea.

Entre sus medidas obliga a los Estados miembros regular las *deepfakes* únicamente de naturaleza sexual o pornográfica. Para ello establece dos argumentos: la fácil, rápida y amplia distribución que se puede producir y la naturaleza íntima de esa imagen, vídeo o audio en la que se han de representar actividades sexuales explícitas o partes íntimas de una persona sin su consentimiento.

En ese sentido, la Directiva destaca que el delito que han de tipificar los Estados miembros debe recoger diversos elementos objetivos:

i) los contenidos *deepfakes* pueden representarse mediante múltiples maneras: como imagen, fotografías o vídeos, incluidas las imágenes sexualizadas, las notas de audio y vídeos cortos.

ii) el acto ha de ser accesible al público mediante las TIC, todo ello sin el consentimiento de la víctima, con independencia de que esta haya consentido la generación de dicho material o pueda habérselo transmitido a una persona en particular.

iii) y se han de incorporar como conductas típicas en el tipo penal: la producción, manipulación y alteración no consentidas a través de técnicas de IA (como, por ejemplo, la edición de imágenes), y que supongan la creación de un material simulado en el que parezca que una persona está llevando a cabo actividades sexuales y ese material se haga accesible posteriormente al público mediante las TIC sin el consentimiento de la víctima.

En el ámbito nacional, existe un interés en tipificar estas conductas en el Código Penal y a tal efecto se aprobó el Proyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales**,** en el que se pretende introducir el artículo 173 bis CP con el objetivo de proteger a las personas víctimas, sobre todo mujeres, de la creación de *deepfakes* que atenten contra su integridad moral, lo que se desarrolla posteriormente en este trabajo[13].

[13] *Vid. Infra*, Capítulo III.

CAPÍTULO II

LAS *DEEPFAKES* SEXUALES COMO NUEVA FORMA DE VIOLENCIA CONTRA LAS MUJERES

1. INTRODUCCIÓN Y CONCEPTO

La era tecnológica actual ha propiciado que además del espacio analógico exista el ciberespacio. Dicho planteamiento resulta relevante pues se trata de dos lugares diferenciados, aunque con la capacidad de relacionarse e influir uno respecto del otro. Por ello, algunos de los hechos que existen en el mundo *offline* se extrapolan al virtual y, en consecuencia, cada vez es más común emplear las tecnologías de la información y comunicación (TIC) para ejercer violencia sobre las mujeres[14].

En ese sentido, el legislador español ha tipificado varias conductas delictivas en las que se utilizan las TIC como medio comisivo. Algunas de ellas son la difusión no consentida de imágenes (conocido como delito de *sexting de tercero*); el acoso predatorio (o delito de *stalking*); contactar mediante las TIC con un menor con propósitos sexuales (delito de *child grooming*); la promoción, fomento o incitación mediante medios tecnológicos al suicidio, a las autolesiones, al consumo de productos peligrosos a menores o personas con discapacidad necesitadas de especial protección o a cometer determinados delitos de carácter sexual contra ellos, entre otros muchos hechos delictivos que, debido a la extensión del presente trabajo no resulta conveniente incidir.

Precisamente ese progreso tecnológico junto al auge de la Inteligencia Artificial ha generado que las técnicas de aprendizaje profundo *deep learning* se conviertan en una herramienta idónea para ejercer violencia sobre las mujeres de muchas maneras. Como ya he adelantado, entre ellas se encuentra la elaboración de contenidos audiovisuales hiperrealistas pero falsos que atenten o pongan en peligro la integridad moral de las mujeres.

[14] Según el Informe "Mujeres jóvenes y acoso en redes sociales" elaborado por el Instituto de las Mujeres, el 80% de las mujeres ha sufrido alguna situación de acoso en las redes sociales. (Disponible en: https://www.inmujeres.gob.es/actualidad/noticias/2022/MAYO/acosorrss.htm). (Fecha de recuperación: 28 de mayo de 2025).

Si bien es cierto que la manipulación de las fotografías no es un fenómeno actual, realmente lo innovador consiste en llevarlo a cabo mediante técnicas de IA generativa (*deep learning*), y ello ocasiona que se generen contenidos tan realistas que incluso resulte complejo diferenciar la autenticidad y la falsedad de los materiales elaborados (Montesinos García, 2024:577).

Por ello, tal y como se ha analizado previamente, mediante la tecnología *deepfake*, se puede elaborar contenido pornográfico sin consentimiento, dado que estas herramientas tecnológicas presentan la capacidad de generar materiales en los que se representen a personas realizando actos, ya sean de carácter sexual o no, que puedan atentar contra su dignidad. Asimismo, mediante las ultrafalsificaciones se consigue una sofisticación considerable del material manipulado y por lo tanto, una mejora en la comisión de diversos delitos.

Debido a ello y dadas sus características, podría entenderse que la elaboración de *deepfakes* que atenten contra la integridad moral puede ser una nueva forma de ejercer ciberviolencia contra las niñas y las mujeres.

Según el Instituto Europeo de Igualdad de Género (*European Institute Gender Equality*) la ciberviolencia de género consiste en aquella *"violencia de género que se perpetra a través de las comunicaciones electrónicas e Internet. Si bien, la civerviolencia puede afectar tanto a mujeres como a hombres, las mujeres y las niñas experimentan formas diferentes y más traumáticas de ciberviolencia. Existen diversas formas de ciberviolencia contra mujeres y niñas, que incluyen, entre otras, el ciberacoso, la pornografía no consensual (o "pornografía de venganza"), insultos basados en el género, el discurso de odio y el acoso, "vergüenza", pornografía no solicitada, "sextorsión", amenazas de violación, amenazas de muerte y redes de trata y tráfico de seres humanos facilitado electrónicamente. La ciberviolencia no es un fenómeno separado de la violencia del "mundo real", ya que a menudo sigue los mismos patrones que la violencia fuera del 'en línea'".*

En esa línea, el Observatorio Nacional de Tecnología y Sociedad (2022:5) dispone que este tipo de violencia no se ha de tratar como un fenómeno aislado sino como una prolongación en el mundo virtual de aquella violencia que pueden sufrir las mujeres en el mundo analógico. Incluso puede surgir en el ciberespacio y trasladarse al espacio físico.

Recordemos que la causa tanto de la violencia *online* como *offline* es la desigualdad estructural de la que surge la voluntad de sometimiento y control de los hombres sobre las mujeres. En otras palabras, se manifiesta como un tipo de violencia de control en la que el sujeto activo aprovecha todas las facilidades que ofrecen los medios tecnológicos, provocando que las conductas sean más lesivas y complejas de evitar, prevenir y castigar (Lloria García, 2024:182).

A su vez, atendiendo a las características de los contenidos *deepfake*, existen dos escenarios en los que pueden convertirse en una nueva manifestación de violencia sobre la mujer. Por un lado, encontramos los supuestos en los que la víctima objeto de dichas manipulaciones es una mujer con relevancia política, social o una celebridad. En estos casos, generalmente la creación y difusión de contenidos audiovisuales manipulados se encuentra estrechamente vinculada con la manosfera y el *gendertrolling* (Lloria García, 2025:83 y Barrientos-Báez, Piñeiro Otero & Portó Renó, 2024:3). Por otro lado, encontramos aquellas situaciones en las que se altera el material de una mujer anónima, normalmente esta conducta se vincula con el concepto de "pornografía de la venganza" y el delito de sextorsión. Ambos escenarios pueden producirse con facilidad dado que existen aplicaciones que permite la creación *deepfakes* sin la necesidad de ostentar elevados conocimientos técnicos ni informáticos (Barrientos-Báez, Piñeiro Otero & Portó Renó, 2024:3).

Asimismo, estas conductas conllevan múltiples riesgos, no solamente para la persona implicada sino también para el individuo del que se ha utilizado su cuerpo. Por lo tanto, mediante la creación y difusión de *deepfakes*, se suplanta la identidad tanto de la persona cuyo rostro se superpone sobre un contenido pornográfico, como de la mujer cuyo cuerpo se utiliza (McGlynn & Toparlak, 2025:207 y Cole, 2024). La cuestión, para el castigo, es si se consigue identificar o no a la persona.

Por su parte, existen ciertas diferencias cuando participan hombres en los contenidos simulados en lugar de mujeres. Normalmente, las *deepfakes* en las que se involucra a políticos, como sucedió con Barack Obama en el programa *BuzzFeed* presentan como objetivo entretener y mostrar el poder tecnológico de las ultrafalsificaciones. Sin embargo, cuando se crean y se difunden *deepfakes*

pornográficas[15] en las que se encuentran involucradas las mujeres, el propósito es distinto, ya que mediante estas creaciones digitales se degrada, se cosifica y se humilla a las mujeres (Burkell & Gosse, 2020:499).

No obstante, como cualquier otro delito de carácter tecnológico, la generación de *deepfakes* implica dificultades en la perseguibilidad, por varias razones. Una de las más relevantes es el anonimato (Almenar Pineda, 2018:27), ya que la mayoría de delitos de difusión y creación de ultrafalsificaciones sexuales se cometen por usuarios que utilizan un pseudónimo. Asimismo, otra de las cuestiones que provocan esa complejidad es que existe una gran demanda en el consumo de pornografía y además muchas veces la distribución puede llevarse a cabo mediante automatización (Maddocks, 2020:420). Por ello, resulta razonable considerar que las *deepfakes* se han convertido en una herramienta delictiva idónea de fácil comisión y difícil persecución.

También debemos tener en cuenta que la IA generativa puede emplearse como un recurso para la investigación judicial y policial para combatir la ciberdelincuencia de género. En particular, la utilización de la IA en estas actividades ha destacado en los ámbitos de la predicción, prevención y actuación policial (Dolz Lago, 2022:21-22 y González Pulido, 2023:170). Asimismo, también puede utilizarse en el campo de la investigación de carácter criminal mediante la identificación biométrica, la realidad aumentada e incluso para detectar posibles amenazas en línea (González Pulido, 2023:170).

En definitiva, en el presente capítulo a fin de detectar de qué manera y de qué forma las ultrafalsificaciones pueden ser un instrumento digital idóneo para ejercer violencia sobre las mujeres, en primer lugar, se analizarán las similitudes y diferencias que presentan las *deepfakes* con otros ciberdelitos de género. En segundo lugar, se estudiará de qué forma pueden emplearse como una herramienta para ejercer control y dominación. Para ello, diferenciaré cómo las *deepfakes* afectan a aquellas mujeres que ocupan un cargo político o presentan relevancia social, y su vinculación con la manosfera de los casos en

[15] MCGLYNN (2025:205) señala que, tal y como han fundamentado autoras como, Mary Anne Franks y Danielle Keats Citron, prefiere denominar dicho término como "falsificaciones sexuales digitales" (*sexual digital forgeries*) debido a que entiende que capta mejor su naturaleza y el daño que pueden generar estos materiales.

los que se utilizan como chantaje emocional con la finalidad de amenazar, intimidar o acosar a las mujeres.

2. LAS *DEEPFAKES* SEXUALES Y OTROS CIBERDELITOS DE GÉNERO: ANÁLISIS Y PERSPECTIVA

El desarrollo tecnológico ha provocado que además de cometer conductas delictivas en el mundo analógico, se atente o se pongan en peligro bienes jurídicos en el ciberespacio[16] empleando como medios comisivos las TIC. Por lo tanto, eso incluye que existan tipos penales que no sean neutros desde el punto de vista de género, dando lugar a lo que se denomina ciberviolencia de género[17].

En ese sentido, el Instituto Europeo de la igualdad de género (2017:1) en su informe "*La ciberviolencia contra mujeres y niñas"* señala que el uso generalizado de las redes sociales, la veloz propagación de la información móvil y el alcance de Internet cada vez más amplio ha ocasionado que la violencia digital contra mujeres y niñas sea un problema significativo que aumenta progresivamente y provoca unas grandes consecuencias económicas y sociales. Sin embargo, en el ámbito comunitario, no se ha logrado conceptualizar de manera explícita, posiblemente debido a la escasez de datos y de estudios relacionados con esta cuestión. De igual modo, el Convenio de Estambul tampoco recoge estrictamente la violencia digital de género, aunque se puede incluir dado que en su artículo segundo se refiere que se entenderá como violencia de género todas las formas de violencia contra las mujeres[18], por lo que en ese caso, también podemos incluir las violencias que utilizan de las TIC (Lloria García, 2025: 73-74).

---

[16] La Real Academia de la lengua española lo define como *"el ámbito virtual creado por medios informáticos"* (Disponible en: https://dle.rae.es/ciberespacio). Asimismo, el Diccionario panhispánico del español jurídico concreta el término señalando que es un *"término acuñado por William Gibson en su novela Neuromante (cyberspace) y es referido al mundo real o artificial generado mediante la conexión a internet"* (Disponible en: https://dpej.rae.es/lema/ciberespacio).
En ese sentido, Santana-Soriano & Báez Vizcaíno proponen definir el término como un espacio artificial y ficcional emergente donde se producen relaciones sociales entre individuos, máquinas y organizaciones, y además no existe de forma autónoma, ya que depende del conjunto de programas informáticos y de equipos que lo posibilitan (Santana Soriano & Báez Vizcaíno 2022:49).

[17] En el ámbito internacional, la primera resolución que relaciona la violencia digital con la violencia contra las mujeres es la Sentencia del TEDH, Caso Buturuga vs Rumanía, 11 de febrero de 2020. A pesar de que resulta una sentencia relevante, en ella se enlazan los dos términos de una forma simple, centrándose únicamente en cuestiones de intimidad (Lloria García, 2024:175).

[18] El párrafo primero del artículo 2 del Convenio de Estambul señala que "*el presente Convenio se aplicará a todas las formas de violencia contra las mujeres, incluida la violencia doméstica, que afecta a las mujeres de manera desproporcionada"*.

La Directiva (UE) 2024/1385 sobre la lucha contra la violencia contra las mujeres y violencia doméstica se ocupa por primera vez en un texto normativo europeo de la violencia digital de género, instando a los Estados miembros a tipificar diversos ciberdelitos, pero no recoge una definición propia de lo que se entiende por violencia digital de género[19] (Ministerio de Igualdad, 2025:79). En particular su apartado 17 alude que *"la ciberviolencia se dirige y afecta especialmente a las mujeres políticas, periodistas y defensoras de los derechos humanos*".

Precisamente, mediante la Directiva se corrobora que este tipo de violencia puede tener el efecto de silenciar y obstaculizar la participación social de las mujeres en condiciones de igualdad con los hombres. Además, el texto comunitario también señala que la ciberviolencia afecta desproporcionadamente a las mujeres y niñas en ámbitos educativos, como son las escuelas y las universidades, y ello conlleva que se produzcan consecuencias perjudiciales para que continúen con su educación y su salud mental, causa exclusión social, ansiedad y tendencia a la autolesión e incluso en casos muy extremos puede desembocar en el suicidio.

Por ese motivo, resulta conveniente entender que la violencia contra las mujeres por razón de género es una cuestión que no solamente se reduce a las relaciones de pareja como indica la Ley Orgánica 1/2004, de 28 de diciembre, de Medidas de Protección Integral contra la Violencia de Género (Lloria García, 2025:71). Prueba de ello es que la desigualdad estructural que existe en el mundo analógico se extrapola al mundo virtual con unas consecuencias y características específicas que ostentan los delitos tecnológicos. En un sentido amplio y a modo conceptual, Lloria García (2020:67) los define como aquellos que o tienen por objeto de ataque un sistema informático, o se generan en el ciberespacio o emplean la tecnología como medio comisivo. En otras palabras, lo constituyen aquellos ilícitos penales en los que interviene algún elemento de la tecnología, sin constituir necesariamente la elaboración de una categoría autónoma de delito. A su vez, las características comunes de los delitos tecnológicos (anonimato, viralidad y permanencia) conllevan un incremento del injusto respecto de los ataques a bienes jurídicos que se producen en el mundo analógico (Lloria García, 2023: 201-202).

[19] A diferencia de la Directiva 2024/1385, la propuesta de la Directiva, sí recogía una definición de ciberviolencia en su apartado cuarto. Señalaba que se entendía como tal *"todo acto de violencia regulado por la presente Directiva cometido, asistido o agravado en parte o en su totalidad mediante el uso de las tecnologías de la información y de las comunicaciones".*

En ese sentido, a efectos de conceptualizar la creación y difusión de ultrafalsificaciones que atenten o que pongan en peligro la integridad moral de las mujeres, resulta conveniente analizar las posibles similitudes y diferencias del delito de *deepfakes* con otros ciberdelitos que no son neutros desde el punto de vista del género.

En primer lugar, la Directiva (UE) 2024/1385 pretende que los Estados miembros tipifiquen como delito la difusión no consentida de imágenes íntimas, también denominado delito de *sexting* de tercero *o sexting* ajeno. Este ilícito penal, al igual que sucede con el delito de creación y difusión de contenidos audiovisuales humillantes, es catalogado como un ciberataque réplica. En España, desde la reforma del año 2015, ya se encuentra recogido como delito contra la intimidad en el artículo 197.7 del Código Penal en el que se regula como conducta delictiva la difusión, revelación o cesión a terceros sin consentimiento imágenes o grabaciones audiovisuales íntimas que han sido obtenidas con consentimiento.

Por lo tanto, la ausencia de consentimiento recae en esa difusión, cesión o revelación y no en la obtención de dichas imágenes (Lloria García, 2020:135). Ello supone exista una clara diferencia con el delito de creación y difusión de contenidos manipulados y es que, en este caso, los materiales corresponden a personas reales y, por esta razón, se atenta o se pone en peligro la intimidad de una persona y no su integridad moral[20]. No obstante, la Directiva en su artículo 5 regula conjuntamente las imágenes reales (en el apartado 5a) y las imágenes manipuladas (en el apartado 5b) bajo el parámetro "difusión no consentida de material íntimo o manipulado"[21].

[20] La Directiva (UE) 2024/1385, en su apartado 19, insta a los estados miembros a tipificar como ilícito penal la difusión de contenidos audiovisuales íntimos sin consentimiento (el delito de sexting tipificado en el art. 197.7 CP). Concretamente, afirma *"la presente Directiva debe incluir todos los tipos de material de este tipo, como imágenes, fotografías y vídeos, incluidas las imágenes sexualizadas, las notas de audio y los vídeos cortos."*

[21] El apartado a) del artículo 5 de la Directiva (UE) 2024/1385 dispone que:

> "*hacer accesible al público, mediante tecnologías de la información y de las comunicaciones (TIC), imágenes, vídeos o materiales similares que representen actividades sexualmente explícitas o las partes íntimas de una persona sin su consentimiento, cuando sea probable que tal conducta cause graves daños a esa persona"*. Por otro lado, el apartado b) de ese mismo artículo regula *"producir, manipular o alterar y, posteriormente, hacer accesible al público, mediante TIC, imágenes, vídeos o materiales similares, haciendo que parezca que una persona está practicando actividades sexualmente explícitas, sin el consentimiento de dicha persona, cuando sea probable que tal conducta cause graves daños a esa persona"*.

Por otro lado, la Directiva 2024/1385, en su artículo 6[22], también obliga a los estados miembros a regular el delito de *stalking* o acoso predatorio (*ciberstalking*). Sin embargo, en nuestro país ya se encuentra tipificado en el artículo 172 ter del Código Penal desde la aprobación de la Ley Orgánica 1/2015, de 30 de marzo, de modificación del Código Penal. En particular, se castiga el acoso de una persona hacia otro individuo de forma insistente y reiterada, sin estar legítimamente autorizado. Esta situación puede producirse a través de distintas conductas que alteran el normal desarrollo de la vida cotidiana del sujeto pasivo:

i) vigilar, perseguir o buscar su cercanía física

ii) intentar establecer comunicación mediante cualquier medio (incluido las TIC) o a través de terceras personas

iii) usar indebidamente sus datos personales para adquirir productos o mercancías, o contratar servicios, o conseguir que terceros se pongan en contacto

iv) atentar contra su libertad o contra su patrimonio o contra la libertad o patrimonio de personas próximas a la víctima.

De tal modo, un elemento del tipo es la ausencia de legitimación para llevar a cabo las conductas anteriores y, además, se requiere la producción de un resultado que consiste en alterar el normal desarrollo de la vida cotidiana de la víctima a través de dichos comportamientos (Ministerio de Igualdad, 2024:93). En tal caso, y atendiendo a la era tecnológica actual, la Directiva (UE) 2024/1385 en su apartado 21 hace hincapié en el control y la vigilancia a través del uso de las TIC[23].

---

[22] El artículo 6 de la Directiva (UE) 2024/1385 regula el ciberacecho y desarrolla que *"los Estados miembros garantizarán que sea punible como delito la conducta intencionada de someter reiterada o continuamente a otra persona a vigilancia, sin el consentimiento de esa persona o una autorización legal para hacerlo, mediante TIC, a fin de rastrear u observar los movimientos y actividades de dicha persona, cuando sea probable que tal conducta cause graves daños a esa persona".*

[23] En particular, el apartado 21 de la Directiva (UE) 2024/1385 señala que

> "*El ciberacecho es una forma moderna de violencia que a menudo se comete contra familiares o personas que viven en el mismo hogar que el autor, aunque también lo cometen exparejas o conocidos. Normalmente, el autor hace un uso indebido de la tecnología para intensificar un comportamiento coactivo y controlador, la manipulación y la vigilancia, incrementando con ello el miedo y la ansiedad de la víctima y su aislamiento gradual de amigos y familiares y del trabajo. Por lo tanto, deben establecerse normas mínimas sobre el ciberacecho. El delito de ciberacecho debe comprender la vigilancia reiterada y continua de la víctima sin su consentimiento o sin autorización legal mediante TIC. Esto podría conseguirse mediante el tratamiento de los datos personales de la víctima, por ejemplo a través de una usurpación de identidad, del robo de contraseñas, del pirateo de los dispositivos de la víctima, de la activación furtiva de software que registre las pulsaciones que se realizan en el teclado para así obtener acceso a espacios privados de*

Y no parece existir obstáculo para considerar que este delito se convierta en tecnológico si, por ejemplo, el autor, mediante el uso indebido de los datos personales de la víctima, adquiere productos o mercancías, o contrata servicios, o hace que terceras personas se pongan en contacto con ella. En tal caso, el sujeto activo puede cometer esta conducta empleando IA generativa o *deep learning*. Imaginemos que a través de la tecnología *deepfake* consigue acceder al teléfono móvil y obtener así datos personales o suplante su identidad y haga que otras personas contacten con el sujeto pasivo.

Pero más allá de los delitos indicados en la Directiva (UE) 2024/1385, es posible perpetrar otros ilícitos empleando técnicas de aprendizaje profundo o *deep learning*. Es el caso del delito de *child grooming*[24] tipificado en el artículo 183 ter del Código Penal, pues puede suceder que una persona cree un contenido audiovisual simulado y lo comparta con una menor a efectos de engañarlo y cometer, de este modo, un delito contra su libertad sexual. Lo mismo podría ocurrir, aunque sin constituir delito de *grooming*, con una víctima mayor de edad. Imaginemos que un sujeto está chateando y hablando con una persona haciéndose pasar por otra empleando una ultrafalsificación y consigue concertar una cita en el mundo analógico con el propósito de cometer un delito contra su libertad sexual.

No obstante, a modo de reflexión y siguiendo a Lloria García (2020:64), no se ha de demonizar la herramienta tecnológica sino todo lo contrario. Las TIC son positivas y constituyen un avance que los ciudadanos puedan usarlas en el día a día dado que facilita la comunicación, el acceso a la información, el estudio, la investigación, entre otras cosas. Precisamente, hay que enfatizar y criminalizar las actuaciones delictivas que se emplean usando esas herramientas y no el término en sí.

la víctima, a través de la instalación de aplicaciones de geolocalización, incluidos programas informáticos de acecho (stalkerware), o robando dispositivos de la víctima".

[24] Este delito se encuentra tipificado en el artículo 183 ter del Código Penal y consiste en contactar con un menor de dieciséis años mediante Internet, del teléfono o de cualquier otra tecnología de la información y la comunicación y proponerle un encuentro a efectos de cometer algunas de las conductas reguladas en el artículo 181 (delito de agresión sexual a menores) y 189 (delito de pornografía infantil).

3. LAS ULTRAFALSIFICACIONES COMO NUEVO MÉTODO DE EJERCER CONTROL Y DOMINACIÓN SOBRE LA MUJER

Las características y el alcance de crear y difundir contenidos simulados mediante sistemas de IA generativa, que atenten o pongan en peligro la integridad moral de las personas, se ha constituido en una conducta potencialmente peligrosa que precisa ser analizada y estudiada, con la finalidad de desarrollar de qué manera se puede ejercer control y dominación, especialmente, sobre las niñas y las mujeres.

En ese sentido, según Okolie (2023:3-5) los perpetradores de la violencia sexual basada en imágenes y por tanto, de *deepfakes*, pueden presentar las siguientes motivaciones:

i) es posible su comisión por mero placer sexual
ii) para acosar a una persona como un ejercicio de poder, en este caso, por ejemplo, en las relaciones de pareja a efectos de amenazar, chantajear y abusar de las mujeres
iii) para evitar el consentimiento
iv) como venganza, esta situación a mi juicio la podemos relacionar con la pornografía de la venganza (*revenge porn)* en la que el perpetrador emplea estos contenidos a efectos de dañar su reputación o ejercer poder sobre ella
v) como prueba de masculinidad, en tal caso, a mi entender se podría vincular con el término manosfera
vi) para ejercer sextorsión sobre la víctima
vii) se crea o se difunden estos contenidos manipulados a efectos de dañar la reputación social del sujeto pasivo, en este caso, de la mujer. Este último motivo podemos relacionarlo también con la violencia que se ejerce sobre las candidatas a las elecciones o las políticas, en los que los usuarios que generalmente componen la manosfera pueden crear y difundir contenidos *deepfakes* con el objetivo de menospreciarlas por el mero hecho de ser mujeres.

En el caso de las *deepfakes* sexuales pueden ser entendidas como una manifestación de ejercer violencia sobre las mujeres, debido a que el rostro de la víctima se trata como un simple recurso de carácter tecnológico que se puede ir editando mediante técnicas de

IA en cuerpos sexuales ajenos o inventados. Por ende, las *deepfakes* perpetúan la idea de que las mujeres son tratadas como simples objetos sexualizados (Van der Nagel, 2020:426). Asimismo, también mediante la creación y difusión de contenidos manipulados se elimina la capacidad de acción y autonomía tanto de las mujeres que se superpone su rostro en el video como de aquellas que se utiliza su cuerpo.

En ese sentido, atendiendo a la gran cantidad de aplicaciones que existen en Internet, así como de imágenes y videos en línea, cualquier mujer puede ser víctima de *deepfakes* sexual. Y es que a través de los contenidos *deepfakes* es posible hacer que cualquier mujer aparezca en un vídeo pornográfico, simplemente se requiere un *corpus* suficiente de imágenes del rostro, un vídeo pornográfico idóneo en el que colocar el rostro de la mujer y algunos conocimientos técnicos (Gosse & Burkell, 2020:498). En definitiva, todo ello evidencia la necesidad de valorar la adecuación del castigo.

3.1. Las ultrafalsificaciones como herramienta para cometer violencia contra las mujeres en política. Influencia de la manosfera en la creación y difusión de deepfakes de contenido sexual. Especial referencia al término *gendertrolling*

El ciberespacio es una herramienta idónea para dar visibilidad a las mujeres, puesto que, en muchas ocasiones, puede llegar a actuar como un enorme altavoz. No obstante, también ha sido (y continúa siendo) un espacio en el que se ejerce violencia sobre ellas por el mero hecho de serlo, aunque con un impacto todavía mayor debido a su capacidad de alcanzar a un amplio número de personas. Y es que, tal como expone Villar-Aguilés & Pecourt Gracia (2021:33) la ciberviolencia contra las mujeres es una corriente profunda de Internet que se encuentra relacionada con el andocentrismo que lo caracteriza desde sus orígenes.

La creación de grupos misóginos surge con los movimientos por la liberación masculina de los años setenta y el movimiento de los derechos de los padres en los años noventa, como con la *National Coalition for Men* creada en el año 1977 (actualmente todavía activa) (Lacalle, Gómez Morales y Vicent-Ibáñez, 2023:3). Por tanto, resulta lógico pensar que a causa del cambio sustancial en la comunicación que ha ocasionado el desarrollo tecnológico, estos grupos antifeministas hayan trasladado también su discurso a este espacio virtual.

La manosfera se puede definir como un conjunto de espacios virtuales diversos que albergan a múltiples movimientos masculinistas centrados en la difusión de discursos antifeministas y misóginos (García-Mingo, Díaz Fernández & Tomás-Forte aludiendo a Ging & Siapera, 2022:3). No obstante, la misoginia en el mundo virtual no es un fenómeno novedoso. Es más, la Declaración de Beijing del año 1995 ya intentaba otorgar medidas destinadas a proteger a las mujeres en el ciberespacio (Ging & Siapera, 2018:517).

Alguno de los subgrupos más populares y significativos son los siguientes. En primer lugar, los MRA (*Men Rights Activists*) que podríamos traducirlos en castellano como los Activistas por los Derechos de los Hombres. En particular, los MRA pueden representar el lado más conservador de los MLM (*Men´s Liberation Movement*) puesto que se originaron tras su desaparición. Sus objetivos primordiales son la protección de la familia y las condiciones de los divorcios (en particular, aluden a la precariedad económica que provocan los divorcios a los hombres y el carácter de las mujeres que utilizan a los hijos con la finalidad de obtener sentencias más favorables para ellas). Los MGTOW (*Men Going their Own Way*) en castellano significa los hombres que siguen su propio camino. Como ocurre con los MRA, este grupo de personas son hombres que se hallan descontentos con la sociedad dominada por el ginecentrismo. Sin embargo, no intentan combatir o luchar contra las mujeres sino que intentan evitarlas. Entre sus actitudes se encuentra la negativa a mantener relaciones sexuales con mujeres o defender la monogamia a efectos de eludir la interposición de denuncias por presuntos delitos contra la libertad sexual. Otro grupo significativo son los PUA (*Pick Up Artits)*, no distan mucho de los MGTOW pero en este caso no las evitan sino que intentan engañar a las mujeres con el objetivo de seducirlas. Y finalmente, los Incels (*Involuntary Celibates*) en castellano se traduce como Celibato involuntario. Este conjunto es similar a los MGTOW puesto que tienen en común los problemas para relacionarse con mujeres (Lacalle, Gómez-Morales & Vicent-Ibáñez, 2023:4-5).

A mi juicio y dadas las características tanto personales como de los medios, estos grupos o movimientos pueden emplear las ultrafalsifcaciones como herramienta con la finalidad de denigrar a las mujeres. Por ello, tal como expone Barrientos-Báez, Piñeiro Otero & Porto Renó (2024:3), las *deepfakes* se han de entender como parte del

gendertrolling, término que utiliza Karla Mantilla para describir un tipo de troleo específico dirigido contra las mujeres.

En ese sentido Karla Mantilla defiende que (2013:563) *"la aparición de las comunidades en línea ha dado lugar a nuevas formas de sexismo virulento que requieren un análisis más profundo. Si bien el comportamiento agresivo en línea (el trolling) está presente en todas las culturas de internet, vale la pena nombrar específicamente las variantes misóginas como "gendertrolling". El gendertrolling, sostengo, tiene varios componentes que lo distinguen de otras formas más genéricas de trolling en línea".*

Este concepto presenta una serie de características. En primer lugar, se trata de un troleo concreto de naturaleza misógina, relacionado con las masculinidades tóxicas (Barrientos-Báez, Piñeiro Otero & Porto Renó 2024:4 y Lloria García, 2025:85). En ese sentido, los usuarios que pueden ser tanto personas como perfiles *bot* que emplean un lenguaje muy agresivo y grosero e, incluso, realizan amenazas de gran intensidad. Existen diversos tipos de manifestaciones dirigidas hacia las mujeres: pueden difundir contenidos pornográficos, imágenes de carácter machista ya sean manipuladas como reales incluso coacciones o amenazas que devengan en algún tipo de situación de acoso (Lloria García, 2025:85). También se encuentran la creación y difusión de mensajes o memes machistas, acoso, intimidación, amenazas de muerte o violación (Barrientos-Báez, Piñeiro Otero & Porto Renó 2024:4). Y actualmente, con los avances de la IA, se podría añadir la creación y difusión no consentida de *deepfakes* pornográficas o vejatorias que vulneren o pongan en peligro la integridad moral de las mujeres.

En segundo lugar, los usuarios no se hallan en una única plataforma virtual, lo hacen en varias e incluso en la vida real (Mantilla, 2013:565) y en tercer lugar, una característica única del *gendertrollling* es que casi siempre surge como respuesta a mujeres que denuncian alguna situación sexista (Mantilla, 2013:565) o que tienen una importante proyección pública o posiciones políticas preeminentes. En definitiva, el troleo de género se caracteriza por ser extremadamente misógino, por la participación de múltiples individuos (normalmente orquestadas), por su permanencia en el tiempo y su fuerte lesividad (Barrientos-Báez, Piñeiro Otero y Porto Renó 2024:3).

Por todo ello, coincido con lo que expone García-Mingo, Díaz Fernández & Tomás-Forte (2022:5) *"la manosfera funciona como un cobijo para los hombres que buscan*

debatir, negociar marcos de sentido y descargar su frustración sobre cuestiones vinculadas a la igualdad de género y al feminismo". Por ello, a mi juicio, empleando actitudes relacionadas con el troleo de género se intenta impedir (o como mínimo obstaculizar) el aumento o la propagación del feminismo sobre todo en redes sociales o en Internet debido a que este conjunto de hombres lo contempla como una amenaza para su hegemonía masculina.

De hecho, existen varios ejemplos de mujeres con relevancia social y/o política o celebridades que han sido objeto de la creación y difusión en redes sociales de *deepfakes*, normalmente de carácter pornográfico o sexual.

Uno de los supuestos más significativos fue el caso de la periodista de investigación india Rana Ayyub. En abril de 2018 tras ser invitada a medios de comunicación para comentar los aspectos políticos que envolvían la violación de una niña de ocho años en Cachemira, fue víctima de una campaña de desinformación a través de una serie de tuits falsos y manipulados que supuestamente había escrito ella. Sin embargo, lo más significativo es que se difundió un *deepfake* pornográfico (Maddocks, 2020: 415 y Ayyub, 2018), que provocó que ante esta situación Ayyub borrara su cuenta de *Facebook* y tuviera que acudir al hospital con palpitaciones y ansiedad (Ayyub, 2018).

Debido a todo lo sucedido, Rana Ayyub, interpuso una denuncia ante la policía de Nueva Delhi que cerraría el caso en el año 2020 ante la incapacidad de descubrir a los autores de dichas actuaciones. Por ello, dadas las graves circunstancias, fue necesaria la intervención de las Naciones Unidas, lo que impulsó que el gobierno de la India adoptará medidas (Okolie, 2023:7-8). En particular la ONU llamó a una relatora especial sobre el derecho a la libertad de opinión y expresión, Irene Khan y a otra relatora especial sobre los defensores de los derechos humanos, Mary Lawlor, a efectos de examinar la situación de Rana Ayyub. Estas expertas relatoras emitieron un comunicado en el que, entre otras cuestiones, exponían que debido a la ausencia de condena y de una investigación idónea por parte del gobierno indio, así como el acoso que éste había llevado a cabo contra Ayyub se había provocado que se legitimaran falsamente a los agresores y los ataques, poniendo así en riesgo incluso su seguridad (UN News, 2022).

De igual manera, la política de Irlanda del Norte, Cara Hunter, durante el final de la campaña electoral del año 2022, y antes de ser elegida miembro de la asamblea

legislativa por el partido Socialdemócrata y Laborista (SDLP), fue objeto de un *deepfake* pornográfico en el que participaba en un acto sexual oral simulado que circulaba en línea (Okolie, 2023:8). Precisamente, por el momento y la forma de actuar de estos usuarios en la creación y difusión de ultrafalsificaciones, se demuestra que estas actuaciones misóginas presentaban como objetivo intimidar a Hunter (McGlynn & Toparlak, 2025:216), es decir, mediante dichos comportamientos se intentaba silenciar a una mujer de la esfera pública de la sociedad.

Otra mujer política que recientemente ha sido objeto de una *deepfake* pornográfica viral ha sido la política de EE.UU, Alexandria Ocasio-Cortez. En particular, se viralizó en la red social X, una imagen manipulada mediante IA generativa en la que se representaba a ella mientras un individuo le obligaba a poner la boca sobre sus genitales (Lorena O´Neil, 2024).

Si bien es cierto que la mayoría de *deepfakes* comentadas han sido de carácter pornográfico, también existen mujeres políticas víctimas de ultrafalsificaciones de carácter vejatorio o humillante, como es el caso de Nancy Pelosi. En este caso se trataba de una *cheapfak*e en la que se representaba falsamente a la presidenta de la cámara de representantes de los Estados Unidos aparentemente ebria (Diakopoulos y Johnson, 2020:2073). Eran videos editados en los que la presidenta parecía que hablaba de manera errática, con dificultades de dicción y con la voz afectada. Ambos vídeos fueron compartidos por el presidente Donald Trump y el ex alcalde de Nueva York, Rudi Guiliani (Barrientos-Báez, Piñeiro Otero & Porto Renó 2024:7) que otorgaban a esos vídeos la apariencia de veracidad y contribuían a una campaña de desinformación contra Nancy Pelosi.

Por todo ello, las *deepfakes* se han convertido en una nueva herramienta idónea para ejercer violencia contra las mujeres políticas. Recientemente ha aumentado la violencia política por razón de género. Tal como expone el Informe “La violencia política por razones de género en España”, en Latino América han sido más conscientes de esta problemática y existen numerosos avances respecto de ella. Concretamente, existe desde el año 1999 la Asociación de Concejalas Bolivianas (Acobol), que tras diversas acciones impulsaron la elaboración del Primer Proyecto de Ley Contra el Acoso y la Violencia Política en Razón de Género. En el año 2017, la Organización de Estados Americanos (OEA) a través de la Comisión Interamericana de Mujeres (CIM) publicaba la Ley

Modelo Interamericana para Prevenir, Sancionar y Erradicar la Violencia contra las Mujeres en la Vida Política. Desde un punto de vista legislativo, respecto a la legislación específica en Bolivia en el año 2012 se aprobó la Ley 243 contra el Acoso y la Violencia Política hacia las mujeres. Por otro lado, también Panamá en el año 2021, aprobaron la Ley 202 de 2021 que modifica la Ley 82 de 2013, sobre violencia contra la mujer y la Ley 7 de 2018, referente a la prevención de actos discriminatorios. Al mismo tiempo, en Perú se promulgó en el año 2021 la Ley que previene y sanciona el acoso contra las mujeres en la vida política (Ministerio de Igualdad, 2022:14 y 24-25).

Según este mismo Informe (2022:20), se entiende por violencia política por razón de género *"aquella que está motivada, se expresa o sus consecuencias e impactos son determinantes por razón del género".* En ese sentido, dispone que las autoras, Bardall, Bjarnegård y Piscopo, definen el concepto "motivación por razón de género", como la violencia empleada por los agresores para mantener la dominación tradicional masculina en la esfera política, excluyendo de esta manera a todas aquellas identidades que no pertenecen a dicho grupo (Ministerio de Igualdad, 2022:20).

De tal modo, el Informe citado señala que existen autoras como Bardall y Krook que diferencian los términos "la violencia política contra las mujeres" de la "violencia política por razón de género". Para definir a la primera de ellas, el informe hace alusión a las autoras Guadarrama y Aguilar, para definir que consiste en aquella que se genera contra la mujer por el mero hecho de serlo y se reproduce en contextos como la familia, el Estado o la sociedad. El informe dispone que por "violencia política por razón de género" las autoras, Bardall y Krook, entienden que bajo dicho término se encuentra la incorporación de otras categorías sociales como puede ser la raza, el origen étnico, la clase, la condición social, la religión, las minorías, las identidades, las orientaciones sexuales e incluso la violencia contra los hombres, niñas y niños, y por consiguiente, convierte al término en un concepto más incluyente y extenso (Ministerio de Igualdad, 2022:22).

En ese sentido, a mi entender, podemos describirla como un tipo de violencia que llevan a cabo los hombres con el objetivo de controlar el sistema político (mantener su dominio hegemónico), a efectos de evitar que las mujeres participen en la esfera pública de la sociedad (en este caso como representantes políticas).

Asimismo, esta manifestación de la violencia se puede llevar a cabo en el ciberespaco. De hecho, respecto al ámbito digital, el citado Informe (2022:91) puntualiza que *"la violencia hacia mujeres activas en política toma una nueva dimensión en Internet. Como se ha visto a lo largo de las entrevistas realizadas, los comportamientos agresivos, abusivos y violentos también se ejercen a través de las tecnologías y en especial en las redes sociales, donde toman una mayor escala y visibilidad. Estos comportamientos tienen especial incidencia en fomentar estereotipos tradicionales y los roles de género".*

Desde un punto de vista legislativo, en el ámbito internacional, como hemos citado con anterioridad, la violencia política por razón de género presenta como finalidad eliminar a las mujeres de la esfera política y por lo tanto, a través de estas prácticas se vulneran diversos derechos fundamentales recogidos en múltiples textos internacionales, como, por ejemplo, la Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer (CEDAW) aprobada por la Asamblea General de las Naciones Unidas en el año 1979 que insta a los estados firmantes a erradicar cualquier tipo de violencia contra las mujeres en la vida política y pública del país. De igual manera, el Convenio del Consejo de Europa sobre prevención y lucha contra la violencia contra las mujeres y la violencia doméstica (también conocido como Convenio de Estambul) se encuentra vigente desde el año 2014 y señala que los estados contratantes han de adoptar medidas que combatan la violencia contra las mujeres en cualquier ámbito de la sociedad, ello incluye el ámbito político y parlamentario (Ministerio de Igualdad, 2022:7).

No obstante, desde un punto de vista comunitario, a lo largo de los años se han impulsado diversos textos que abordan esta cuestión. En tal caso, entre otras, se encuentra la Recomendación 288 (2010) sobre igualdad de género en la política, la Recomendación 390 (2016) sobre participación política de las mujeres y la Recomendación 459 Recomendación 449 ambas emitidas en el año 2020 sobre la violencia contra las mujeres en política en el ámbito regional y local (Ministerio de Igualdad, 2022:26-27). Asimismo, recientemente la Directiva (UE) 2024/1385 en su apartado 17[25] expresamente reconoce

[25] El apartado 17 de la Directiva (UE) 2024/1385 señala que "*(...) La ciberviolencia se dirige y afecta especialmente a las mujeres políticas, periodistas y defensoras de los derechos humanos. Los defensores de los derechos humanos son personas, grupos u organizaciones que promueven y protegen los derechos humanos y las libertades fundamentales universalmente reconocidos (...)".*

que, en ocasiones, la ciberviolencia se dirige y afecta principalmente a las mujeres políticas, periodistas y defensora de los derechos humanos.

Por lo tanto, es posible que los hombres que componen la manosfera creen y compartan contenidos misóginos, que tienen como fin humillar, hipersexualizar y silenciar a las mujeres que defienden ideas feministas o que simplemente participan en la esfera pública de la sociedad. Todo ello, como he ido indicando insistentemente, porque ven en peligro su hegemonía masculina y emplean estas herramientas digitales para intimidar a las mujeres y menospreciarlas por el mero hecho de serlo, dotando de contenido así al denominado "troleo de género".

En esa línea el Informe "La violencia política por razón de género en España" basándose en la matriz planteada por las autoras Bardall y Krook, dispone que generalmente las mujeres víctimas de este tipo de violencia suelen ser, mujeres con proyección pública, políticas (que ocupan cargos electos, candidatas o militantes de partidos políticos), o mujeres que ostentan puestos institucionales (tales como funcionaras, altos cargos en entidades administrativas, empleadas del sector público), mujeres profesionales que no tienen un cargo estatal o institucional (como ocurre con las periodistas, representantes de sindicatos, activistas, comunicadoras, etc.) y mujeres anónimas o privadas (Ministerio de Igualdad, 2022:29). En este último supuesto, entiendo que se pueden incluir aquellas que no tienen ningún cargo o proyección política, profesional o social, pero que difunden o comparten algún pensamiento político en cualquier medio de comunicación o que realizan actos que se encuentran polarizados en el ámbito político.

En cambio, además de las mujeres con relevancia política y/o social, otras mujeres célebres en la industria musical o cinematográfica también han sido víctimas de *deepfakes* sobre todo pornográficas o sexuales. Las primeras celebridades en ser objeto de ultrafalsificaciones fueron la cantante Taylor Swift o las actrices Gal Gadot o Scarlett Johansson tras difundirse en el año 2017 en la plataforma *Reddit* unos vídeos manipulados pornográficos por un usuario llamado "deepfake" (Cole, 2017 y Kweilin,2022:650). De la misma forma, recientemente se ha descubierto una red de *Telegram* en Corea del Sur que se dedicaba a crear y difundir contenido *deepfake* sexual o pornográfico de más de 200 cantantes k-pop surcoreanas (Aguayo, 2024 y Martin, 2024).

Por el contrario, y a pesar de que es evidente que los hombres con cargos políticos también han sido víctimas de ultrafalsificaciones, lo cierto es que los contenidos no suelen presentar las mismas características que los que padecen las mujeres. Por ejemplo, hace unos años como hemos desarrollado con anterioridad, la empresa estadounidense de medios de comunicación *BuzzFeed* con la finalidad de advertir al público del auge de la IA o de mostrar cómo se puede manipular un video en línea, difundió un vídeo del expresidente de Estados Unidos Barack Obama insultando a Donald Trump (Vaccari y Chadwick, 2020:1). Otro caso relevante fue el contenido audiovisual manipulado del actor Tom Cruise que se viralizó en *TikTok*. En dicho vídeo se representaba al actor llevando a cabo un truco de magia, riéndose, jugando al golf o cayéndose delante de una tienda (Díaz, 2021).

No obstante, tal y como se ha anunciado, y tras el análisis realizado se comprueba que los contenidos en los que se representan simuladamente a los hombres suelen ser completamente distintos a los que protagonizan las mujeres. En ellos, a los varones generalmente no se les hipersexualiza como sí ocurre con las mujeres, esto es así dado que los materiales presentan un carácter más satírico o artístico.

Es de esta forma, precisamente porque la comunidad *deepfake* refleja una masculinidad tóxica *geek* (Newton & Stanfill, 2020:399), a mi juicio asociada generalmente al conjunto de hombres que componen la manosfera. Por ello, tras lo analizado considero que principalmente, y centrándonos en la creación y difusión de contenidos hiperrealistas creados mediante IA generativa, los usuarios de estas comunidades llevan a cabo estas prácticas a efectos de humillar, cosificar y obstaculizar la presencia femenina en la esfera pública y visible de nuestra sociedad.

3.2. Las *deepfakes* como instrumento para cometer chantaje emocional contra las mujeres. Especial referencia a la pornografía de la venganza y la sextorsión

Las ultrafalsificaciones comenzaron a popularizarse en el año 2017 en la plataforma virtual *Reddit* de una manera muy sencilla. De hecho, tras cobrar fama este tipo de contenidos, un usuario de esa misma plataforma llamado “deepfakeapp” elaboró una aplicación denominada “FakeApp” que permitía a cualquier persona sin experiencia informática poder crear este tipo de contenidos (Beamonte, 2018). No obstante,

actualmente existen bastantes más aplicaciones que permiten su elaboración en pocos minutos. De hecho, según el informe *2023 State of Deepfakes* en tan solo 25 minutos y sin ningún coste, es posible elaborar un vídeo *deepfake* pornográfico de 60 segundos y sin ningún coste (Home Securtity Heroes, 2023). Por lo tanto, no resultaría extraño pensar que cualquier mujer puede llegar a convertirse en víctima de una ultrafalsificación no consentida de carácter sexual o humillante dado que, no es necesario que los autores posean contenidos sexuales de la persona real, simplemente se requiere un conjunto de imágenes del rostro de la persona que se desea manipular, un vídeo pornográfico adecuado en el que superponer el rostro normalmente de la actriz porno por el de la víctima y algunos (no demasiados) conocimientos técnicos y/o informáticos.

A mi juicio, entre las formas de ciberviolencia que se pueden ejercer contra las mujeres se encuentra la "pornovenganza" y la sextorsión. Respecto a la primera de ellas, dadas las circunstancias de la conducta objeto de análisis, así como de sus características, un individuo podría elaborar y difundir un contenido sexual simulado con la finalidad de vengarse, y de esta manera ocasionar un daño y una lesión a la víctima. Imaginemos, por ejemplo, que una mujer ha dejado a su novio, y él decide elaborar y difundir una ultrafalsificación en la que se represente manteniendo una relación simulada con otro hombre. En este caso, el sujeto activo emplearía ese tipo de contenido con el objetivo de vengarse de la víctima con el único propósito de dañar su reputación social o para ejercer poder sobre ella (Okolie, 2023:4). Y es que los actos de la pornografía de la venganza reflejan las desigualdades de género y la cosificación de las mujeres en el espacio virtual, situación que también puede trasladarse a los *deepfakes* sexuales (Barba Arteaga, 2024:235).

De igual manera, mediante las *deepfakes* se puede cometer delito de sextorsión, que consiste en amenazar con distribuir un material íntimo o sexual de una persona con el objetivo de que esta lleve a cabo ciertas exigencias que le imponga el autor (Kweilin,2022:652). Imaginemos, por ejemplo, que una mujer ha decidido romper el vínculo sentimental con su novio y este a modo de chantaje, le comunica que si no vuelve con él difundirá por redes sociales un contenido hiperrealista simulado en el que aparece ella manteniendo una relación sexual con otro varón.

Además, puede ocurrir que la elaboración y difusión de ultrafalsificaciones se empleen para obstaculizar la interposición de denuncias. Por ejemplo, un hombre

presuntamente sujeto activo de un delito puede crear un *deepfake* en el que se cuestione si hubo o no consentimiento o si sucedieron o no los actos denunciados provocando de esta manera que finalmente la víctima desista. Por otro lado, es posible que un individuo conozca que una mujer va a interponer una denuncia, para utilizar la técnica de la contradenuncia se elabora un *deepfake* que le sirva de prueba en el que simule a la mujer cometiendo un ilícito penal contra él (Simó Soler, 2023:501-503).

Por otro lado, además de emplear las *deepfakes* para ejercer venganza o para coaccionar a las víctimas a realizar algo que no desean, se pueden utilizar como medio para cometer otro tipo de ilícitos penales en el mundo analógico. Imaginemos, por ejemplo, un individuo que mediante técnicas de IA generativa de *deep learning* manipula una imagen y se crea un perfil falso o se hace pasar por otra persona a efectos de concertar una cita con una mujer y cometer un delito contra su libertad sexual o contra su vida. En tal caso se estaría empleando la IA generativa como medio comisivo para perpetrar un delito en el espacio analógico. De esta manera, también podría emplear técnicas de IA generativa para clonar el rostro de la víctima y así acceder a su teléfono móvil a efectos de obtener datos íntimos de la persona con el objetivo de compartirlos y revelarlo en redes sociales, cometiendo así un delito de *doxing*.

Por todo ello, existen múltiples formas en las que utilizar *deepfakes* para cometer delitos, por lo que a modo de recopilación tras lo analizado, resulta relevante incidir en que las ultrafalsificaciones no solamente pueden ejercerse con el propósito de obtener placer sexual, sino también al igual que sucede con la pornografía de la venganza y la sextorsión, los hombres abusivos pueden emplear *deepfakes* sexuales no consentidas a efectos de intimidar, aislar y avergonzar a las víctimas, que generalmente son mujeres (Kweilin, 2022:648).

En la actualidad existen varios ejemplos en los que se ha utilizado *deepfakes* contra mujeres anónimas. Un caso significativo tuvo lugar recientemente cuando una joven australiana tras recibir un correo electrónico anónimo con un enlace y darle al *click* se encontró con páginas que contenían múltiples imágenes pornográficas suyas manipuladas mediante IA, en las que su rostro se había superpuesto en escenas de contenido sexual violento. En algunas de ellas se indicaba su nombre completo, así como datos personales como su cuenta de Instagram, su domicilio e incluso su número de teléfono. Las imágenes no solamente la describían en situaciones que atentaban contra su dignidad sino también

estaban acompañadas de amenazas y de encuestas en las que individuos votaban sobre la forma en que querían abusar de ella (en concreto la página web en la que se difundieron esos contenidos se denominaba *"la destrucción de Hannah"* y en ella había más de 600 imágenes falsas). Hannah empezó a investigar el sitio y se percató que había imágenes de otras mujeres. Tras una investigación policial y judicial, se demostró que el autor de dichos contenidos era uno de sus amigos de la universidad que finalmente fue condenado a nueve años de prisión por uso indebido de un servicio de comunicación para acosar, ofender o amenazar. Dicha condena sirvió como precedente para intentar tipificar dichas conductas en Australia (Gómez Isaza, 2025, Turnbull, 2025, Merino, 2025).

Un caso similar sufrió otra joven en Reino Unido, dado que también recibió un correo desde una cuenta anónima con un enlace a una web en la que halló imágenes explícitas y un vídeo *deepfake* pornográfico en el que se representaba falsamente a ella manteniendo relaciones sexuales con varios hombres. A lo largo de las investigaciones se descubrió, al igual que el caso de Hannah, que el autor de dichos contenidos era un amigo de la joven. Finalmente, en agosto de 2021 fue condenado a veinte semanas de prisión, a una suspensión de dos años y a pagar una indemnización a cada una de sus víctimas por tomar imágenes de quince mujeres (incluida la víctima en cuestión) de las redes sociales y subirlas a sitios web de carácter pornográfico (West, 2024).

Por otro lado, en España se produjo el primer caso de *deepfakes* sexuales entre menores de edad en el denominado "Caso Almendralejo" en el que varios menores utilizaron aplicaciones de IA, como *App ClothOff,* para crear imágenes falsas en las que se superponía el rostro de compañeras del instituto en el cuerpo de otras mujeres desnudas y posteriormente se difundieron en grupos de *WhatsApp* (Varona, 2025:229-230).

Desde una perspectiva jurídico penal, en este supuesto y en la hipotética situación que se tipificara finalmente el delito, en el caso de que fueran ultrafalsificaciones de carácter sexual, a mi juicio, se produciría un concurso de normas penales entre el ilícito proyectado y el delito de pornografía infantil virtual regulado en el en el artículo 189 CP[26]. Del mismo modo, a efectos procesales, como algunos de los sujetos activos son menores de dieciochos años y mayores de catorce se aplicaría la Ley Orgánica 5/2000, de 12 de enero, reguladora de la responsabilidad penal de los menores.

[26] En el capítulo III analizaré con mayor profundidad.

Por todo ello, la creación y difusión de *deepfakes* puede tener diversas finalidades con las que se puede ejercer control y dominación sobre las mujeres. En ellas se las cosifica al tratar su cuerpo como un mero objeto sexualizado y en consecuencia, mediante la utilización de esta herramienta se ha puesto de manifiesto que puede emplearse como una nueva forma de ejercer violencia de género en el ciberespacio.

CAPÍTULO III

EL NUEVO ARTÍCULO 173 BIS CP EN EL PROYECTO DE LEY ORGÁNICA PARA LA PROTECCIÓN DE LAS PERSONAS MENORES DE EDAD EN LOS ENTORNOS DIGITALES

1. INTRODUCCIÓN

Tal y como se ha ido exponiendo a lo largo del trabajo resulta evidente que, a causa de los múltiples avances y progresos de las nuevas tecnologías, se ha producido en Europa un aumento considerable de los ciberataques y de la ciberdelinciencia y que son cada vez más sofisticados. Los niños, niñas y adolescentes constituyen un grupo de la sociedad que utiliza todo tipo de herramientas tecnológicas. De hecho, según el Observatorio Nacional de Tecnología en España, en el año 2023, el 94,7% de la población de 10 a 15 años utilizó Internet en los últimos tres meses y el 93,1% usó ordenadores o *tablets*. Asimismo, el 70,6% de los menores disponen de teléfono móvil (Observatorio Nacional de Tecnología y Sociedad, 2023). Por lo tanto, con estos datos, resulta relevante destacar que los menores, al igual que los adultos, pueden encontrarse expuestos a los contenidos que se difunden o se comparten en la Red. La diferencia es que, por razones evolutivas, constituyen un grupo de la sociedad más vulnerable que requiere de una protección especial y específica.

De igual manera, la Agencia Española de Protección de Datos (AEPD) apoyándose en el informe "Mujeres jóvenes y acoso en redes sociales", elaborado por el Instituto de las Mujeres, recoge que el 80% de las mujeres ha sufrido algún tipo de situación de acoso en las redes sociales (Agencia Española de Protección de Datos, 2023). Por lo tanto, como ya he advertido insistentemente, la violencia digital no es una cuestión neutra desde el punto de vista del género y parece que afecta especialmente a las personas jóvenes y a las mujeres de manera particular.

Precisar por ello que, en la Exposición de Motivos del Proyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales se argumenta que la facilidad en el acceso y la globalización de los entornos digitales hacen posible que las personas menores de edad empleen estos medios tecnológicos para el ejercicio de sus derechos fundamentales, tales como el derecho a la información o el derecho a la libertad

de expresión, así como para su participación política, social y cultural en diversos ámbitos, ya sea en el ámbito internacional, nacional o local. Además, destaca que, ocasiones el entorno virtual puede no ser seguro. El ciberespacio puede contener mensajes y contenidos que reproducen estereotipos de género, discriminatorios o violentos o información falsa o sobre hábitos de conducta o consumo perjudiciales o ilegales.

Esto último nos puede recordar al Preámbulo de la Ley Orgánica 8/2021, de 4 de junio, de protección integral a la infancia y la adolescencia frente a la violencia donde se describe que, a causa de la gran alarma social que provocan todas estas conductas, y a efectos de proteger a las niñas, niños y adolescentes frente a cualquier tipo de violencia, se castigará a aquellos sujetos, que mediante las TIC, promuevan, inciten o fomenten el suicidio, las autolesiones, los trastornos alimenticios a personas menores de edad o personas con discapacidad necesitada de especial protección, así como a cometer diversos delitos sexuales contra ellos.

En ese sentido, el legislador entiende que tras los avances de las nuevas tecnologías -especialmente de la IA-, puede resultar justificada la tipificación de nuevas conductas delictivas, con la finalidad de castigar aquellos hechos que lesionen o pongan en peligro bienes jurídicos. Y en esta línea se mueve el Proyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales. Entre sus avances más significativos en materia penal destaca el deseo de crear un nuevo delito, recogido en el artículo 173 bis CP que castiga a quién, con ánimo de atentar contra la integridad moral, difunde, cede o exhibe contenidos audiovisuales generados mediante sistemas informáticos en los que se simulen de manera realista situaciones de contenido sexual o gravemente vejatorias.

Ante esta propuesta, en el presente Capítulo se estudiará el delito proyectado de una manera muy resumida dado que todavía no estamos ante la presencia de un texto definitivo. El análisis se llevará a cabo siguiendo la teoría del delito, lo que conlleva la necesidad de examinar si el lugar que ha elegido el prelegislador resulta sistemáticamente adecuado o no. Para ello y tomando en cuenta a circularidad interpretativa, resulta imprescindible para la determinación del bien jurídico atender también a la redacción típica.

2. BIEN JURÍDICO PROTEGIDO

El análisis de cualquier figura delictiva debe comenzar por el estudio de su bien jurídico protegido o elemento nuclear del tipo. En la ciencia penal y desde la perspectiva del orden social, el bien jurídico protegido se entiende como todo aquello que aparece como un valor positivo y por ello, disfruta de la protección del Derecho, pues cualquier acción típica se encuentra orientada a la puesta en peligro o lesión de bienes jurídicos protegidos (Rodríguez Mourullo,1978: 276).

El proyectado artículo 173 bis CP en el caso de ser tipificado se encontraría ubicado en el Título VII "De las torturas y otros delitos contra la integridad moral" y concretamente, el prelegislador parece que ha decido a través de su redacción, proteger la integridad moral estableciendo que:

> *"Se impondrá la pena de prisión de uno a dos años a quienes, sin autorización de la persona afectada y con ánimo de menoscabar su integridad moral, difundan, exhiban o cedan su imagen corporal o audio de voz generada, modificada o recreada mediante sistemas automatizados, software, algoritmos, inteligencia artificial o cualquier otra tecnología, de modo que parezca real, simulando situaciones de contenido sexual o gravemente vejatorias. La pena será de un año y seis meses a dos años, si la víctima es menor de edad o persona con discapacidad necesitada de especial protección.*
>
> *Se aplicará la pena en su mitad superior si dicho material ultrafalsificado se difunde a través de un medio de comunicación social, por medio de internet o mediante el uso de tecnologías, de modo que aquel se hiciera accesible a un elevado número de personas en el espacio virtual.*
> *Las autoridades judiciales ordenarán la adopción de las medidas necesarias para la retirada de los contenidos a los que se refieren los párrafos anteriores, para la interrupción de los servicios que ofrezcan predominantemente dichos contenidos o para el bloqueo de unos y otros cuando radiquen en el extranjero"*

Esta ubicación y el bien jurídico que parece proteger resulta novedosa. Si atendemos a los antecedentes, se puede afirmar que, por un lado, el grupo parlamentario Plural propuso una enmienda a la Ley Orgánica 4/2023 en la que a través de ella pretendía

introducir la incorporación de un delito de difusión de *deepfakes* pornográficas o vejatorias como delitos contra la intimidad[27]. Sin embargo, tal y como expone Devís Matamoros (2023:2) este planteamiento debe descartarse, pues, si bien es cierto que la ciencia penal ha de adaptarse a la actualidad, no debe permitir introducir nuevas modificaciones que desvirtúen su contenido esencial. Ello se fundamenta en que los materiales son simulados, no representan a una persona real, y por ende, no existe intimidad personal[28] que se tenga que proteger.

Por otro lado, el grupo parlamentario SUMAR presentó una Proposición de Ley Orgánica de "regulación de las simulaciones de imágenes y voces de personas generadas por medio de inteligencia artificial", en la que se pretendía tipificar este delito como un delito de injurias graves, a través de la incorporación del nuevo artículo 208 bis CP[29]. En un contexto similar, Jareño Leal (2024:34-35) defiende que la difusión de materiales sexuales manipulados por IA no ha de entenderse como delitos contra la intimidad, pero sí aboga por tipificarlas a corto plazo como delitos de injurias graves.

Siguiendo a Lloria García (2025:90) no resulta lo más conveniente tipificar estas conductas como delitos contra el honor por diversos motivos:

i) Los contenidos *deepfakes* no son únicamente sexuales o eróticos; existen materiales simulados que, sin revertir el carácter sexual, son también degradantes y humillantes.

ii) Si bien es cierto que, a través de la difusión de las ultrafalsificaciones se puede atentar contra el honor de una persona, dado que se trata de *"conductas vejatorias gravemente humillantes"*, realmente lo que se vulnera es la dignidad en un sentido amplio, provocando situaciones de miedo, angustia y envilecimiento. Por tanto, muchas

[27] Es posible que el Grupo Parlamentario Plural se fundamentara en la Ley Olimpia de Méjico en donde se castiga la difusión no consentida de imágenes reales y simuladas como violencia digital sexual. Dicha idea también la propone la Directiva (UE) 2024/1385 en su artículo 5 (Lloria García, 2025:88).

[28] El derecho fundamental a la intimidad (al igual que el derecho a la propia imagen) se regula en el artículo 18 CE y tiene como finalidad garantizar un espacio propio y reservado de un individuo frente a la acción y el conocimiento de terceros ya sea bien por los poderes públicos o por meros individuos (Doval Pais, 2016: 495-496).

[29] El propuesto artículo 208 bis CP disponía que: *"Igualmente tendrá la consideración de injuria la acción que, sin autorización y con ánimo de menoscabar el honor, fama, dignidad o la propia estimación de una persona, recrease mediante sistemas automatizados, software, algoritmos o inteligencia artificial para la pública difusión su imagen corporal o audio de voz".*

de estas acciones trascienden incluso de la mera afectación a la dignidad por menoscabo de la estimación o fama.

Por el contrario, resulta conveniente incidir en que si el bien jurídico protegido de estas conductas fuera el honor de un individuo, otras conductas como las destinadas a comercializar o cosificar podrían quedarse al margen de este ilícito penal (Almenar Pineda, 2025:118), así como otras intenciones, como puede ser el ánimo lucrativo, debido a que eventualmente es posible que se publiquen en páginas webs pornográficas con la finalidad de obtener más reproducciones y suscriptores (Devís Matamoros, 2023:3).

Esta parece ser la posición del prelegislador, que con esta redacción puede lograr, a través del principio de consunción (artículo 8.3 CP)[30] que la afectación contra el honor se castigue a través del proyectado artículo 173 bis CP por ser este precepto más amplio. De tal modo, la conducta acogería tanto las lesiones contra el honor como las lesiones contra la integridad moral. En esta línea, Devís Matamoros (2023:3) defiende que no solamente se abarcaría la vulneración a la autoestima y la heteroestima, también se castiga la instrumentalización y la cosificación que se genera sobre el sujeto pasivo. Concretamente, el texto prelegislativo en la Exposición de Motivos del Proyecto de Ley Orgánica dispone que:

> *"Técnicamente, se opta por la sanción de la difusión de las ultrafalsificaciones de contenido sexual (las llamadas deepfakes pornográficas) o especialmente vejatorio en sede de delitos contra la integridad moral porque, en virtud del principio de consunción, se abarcarían los supuestos de lesión de la integridad moral y también los ataques contra el honor, pues ha de tomarse en cuenta no solo la afectación a la autoestima y la heteroestima, sino también la cosificación e instrumentalización que se produce sobre el sujeto pasivo, generalmente mujeres y niñas, niños y adolescentes que son tratados como objetos de consumo. También hay que recordar que la motivación para llevar a cabo estas acciones no siempre se identifica con el animus iniuriandi, pues el*

[30] En ocasiones una conducta puede subsumirse en varios tipos delictivos, eso se denomina concurso de leyes o concurso aparente de leyes. Para ello, el Código Penal en su artículo 8 dispone que:

> "*los hechos susceptibles de ser calificados con arreglo a dos o más preceptos de este Código, y no comprendidos en los artículos 73 a 77, se castigarán observando las siguientes reglas: 1.ª El precepto especial se aplicará con preferencia al general; 2.ª El precepto subsidiario se aplicará sólo en defecto del principal, ya se declare expresamente dicha subsidiariedad, ya sea ésta tácitamente deducible; 3.ª El precepto penal más amplio o complejo absorberá a los que castiguen las infracciones consumidas en aquél; 4.ª En defecto de los criterios anteriores, el precepto penal más grave excluirá los que castiguen el hecho con pena menor".*

hecho puede deberse a otras razones como el ánimo de lucro, si dichas imágenes se utilizan en páginas o aplicaciones de contenido pornográfico".

En ese sentido, en atención a la posición sistemática, típica y a la propia redacción del proyectado artículo 173 bis CP, parece evidente que lo que se pretende proteger es la integridad moral del sujeto que se representa simuladamente en dichos contenidos. Como es sabido, generalmente, los autores que generan ultrafalsificaciones superponen el rostro de la víctima en el cuerpo de una actriz porno (normalmente extraído de una película de estas características). Ello nos invita a reflexionar que de dicha actuación se pueden derivar problemas relacionados con la propiedad intelectual, aunque dada la extensión del presente estudio no resulta pertinente desarrollar en profundidad.

El derecho constitucional a la integridad moral se encuentra regulado en el artículo 15 CE[31]. La integridad moral se corresponde con la dignidad humana, pues se trata de una manifestación de la dignidad personal (Grima Lizandra, 2015:291). Particularmente, la STS 223/2009, de 3 de marzo la define como *"la inviolabilidad de las personas como manifestación directa de la dignidad humana"*. En otras palabras, la integridad moral[32] se vincula estrechamente con los conceptos de dignidad humana y el carácter inviolable de las personas. Precisamente castigando las conductas que atentan contra ella, se intenta proteger que todo ser humano sea tratado como un fin en sí mismo y nunca cosificarlo o reducirlo a un mero objeto. Por ello la integridad moral se representa como un valor independiente del derecho a la vida, a la integridad física, al honor o a cualquier manifestación de la libertad (STC 426/2021, de 19 de mayo de 2021, fundamento jurídico 2.5.2).

Siguiendo a Grima Lizandra (2015:292), la filosofía kantiana dispone que la dignidad humana se manifiesta por dos dimensiones[33]. Por un lado, entiende que el ser humano es fin en sí mismo y por ello, no ha de ser tratado como un mero instrumento, es

[31] El artículo 15 CE establece que *"todos tienen derecho a la vida y a la integridad física y moral, sin que, en ningún caso puedan ser sometidos a torturas ni a penas o tratos inhumanos o degradantes".*

[32] La STS 1725/2001, 3 de octubre 2001 (FD 6), dispone la integridad moral como *"un atributo de la persona, como ente dotado de dignidad por el solo hecho de serlo; esto es, como sujeto moral, fin en sí mismo, investido de la capacidad para decidir responsablemente sobre el propio comportamiento. La garantía constitucional de la dignidad, como valor de alta calidad indicada, implica la proscripción de cualquier uso instrumental de un sujeto y de la imposición al mismo de algún menoscabo que no responda a un fin constitucionalmente legítimo y legalmente previsto".*

[33] Según Kant el ser humano como ser racional, existe como fin en sí, contrapone la noción de persona a la noción de cosa, ya que estas últimas ostentan un valor relativo y por ello, es posible emplearlas como medio para conseguir un propósito. En cambio, la persona humana es fin en sí y esto es equivalente a que nunca ninguna persona puede ser tratada como un mero medio (Beade, 2016:25-26).

decir, no puede ser utilizado exclusivamente como una cosa. Por otro lado, estos planteamientos conllevan reconocer la libertad y la autonomía de la voluntad de las personas.

En ese sentido, dado que la integridad moral se ha de entender como un concepto inseparable de la dignidad humana, también posee esas dos dimensiones. De un lado, supone la prohibición de humillaciones o degradaciones de las personas. De otro, conlleva reconocer al ser humano la libertad de autodeterminación y de actuación conforme a lo que previamente ha decidido realizar, esto es reconocer la libertad de voluntad y la libertad de conciencia.

No obstante, para ser considerado delito debe atentar de manera significativa contra la integridad moral -los tratos degradantes ocasionados han de afectar gravemente a la dignidad humana- y para ello, existe un criterio jurisprudencial que analiza la intensidad o la gravedad de la acción. Según la STS 325/2013, de 2 de abril, se ha de examinar la duración, la intensidad, los efectos físicos y mentales, así como las condiciones de la víctima, tales como el estado de salud, el sexo o la edad. Estos criterios, en el supuesto de los contenidos *deepfakes* vejatorios o sexuales, fundamentan la gravedad requerida para anular el principio de intervención mínima y el principio de insignificancia (Lloria García, 2025:90).

Con estos razonamientos estoy en condiciones de afirmar que a mi juicio resulta acertada la interpretación sistemática y típica llevada a cabo por el prelegislador puesto que castigando dichas conductas como atentados contra la integridad moral también se subsumen aquellos que menoscaban el honor a través del ya nombrado principio de consunción. De tal modo, mediante su tipificación, se dispone de una protección completa para los sujetos pasivos de estos ilícitos penales.

3. SUJETOS DEL DELITO

La redacción del Proyecto de Ley Orgánica describe el tipo como un delito común[34], es decir, la conducta delictiva regulada en el proyectado artículo 173 bis CP

[34] Los delitos comunes son aquellos que castigan conductas en las que cualquiera puede ser sujeto activo y por ello suelen comenzar con la expresión "el que" o "los que". En cambio, los delitos especiales requieren ciertas características o cualidades personales, provocando que ya no sea posible que cualquier persona sea

puede ser cometida por cualquier persona con independencia de parentesco, afectividad, cargo u oficio.

Respecto al sujeto pasivo, que es el titular del bien jurídico, puede ser cualquiera pues el hecho de que la conducta generalmente se circunscriba al círculo de los menores no es óbice para que se configure como un delito común. Todo ello con independencia de los tipos agravados que se pueden constituir sobre la base de sujetos especiales.

De hecho, el primer párrafo del artículo proyectado contempla una agravación de la pena en el supuesto de que la persona suplantada digitalmente sea un menor de edad o una persona con discapacidad necesitada de especial protección, al entenderse que se trata de sujetos particularmente vulnerables que requieren de una mayor protección jurídico-penal.

4. CONDUCTA TÍPICA

Por lo que hace a la conducta, esta viene delimitada por los verbos típicos "difundir", "exhibir" y "ceder". Según la RAE el primero de ellos consiste en *"propagar o divulgar conocimientos, noticias, actitudes, costumbres, modas"* (Real Academia Española, s.f.). Por otro lado, "exhibir", conlleva *"manifestar, mostrar en público"* (Real Academia Española, s.f.). Y finalmente, según el mismo diccionario, el concepto "ceder" lo describe como *"dar, transferir o traspasar a alguien una cosa, acción o derecho"* (Real Academia Española, s.f.).

Tal y como expone Lloria García (2020:141), "difundir" conlleva propagar o expandir, es decir, divulgar un contenido masivamente. A través de la difusión -en este caso, la imagen o el audio de voz manipulado-, el contenido se expone públicamente. El término "ceder" supone transmitir o traspasar algo -un contenido, una información- a otra persona. Si lo extrapolamos al ciberespacio se requerirá que se transfiera de un dispositivo a otro -de un usuario a otro usuario-, con el peligro y el riesgo de que ese contenido se viralice. Por último, se puede interpretar que las conductas de exhibición pueden ser comprendidas en la acción "revelar", que implica mostrar algo o alguna cosa a alguien

considerada sujeto activo del delito, en esta categoría de delitos se encuentran los "delitos de propia mano", delitos "especiales en sentido estricto" y delitos "especiales en sentido amplio" (Rodríguez Mourullo, 1978:269).

que no conoce. En el ámbito virtual, exhibir un contenido audiovisual en una red social significa que el usuario que lo muestra continúa disponiéndolo. Sin embargo, el material puede transmitirse dado que cualquier persona puede descargarlo, capturarlo o exhibirlo a otro individuo.

En esta línea, la técnica legislativa que emplea el prelegislador con el delito de difusión sin consentimiento de ultrafalsificaciones pornográficas y gravemente vejatorias recuerda a la utilizada en el delito de *sexting*, regulado en el artículo 197.7 CP[35], debido a que se emplean los mismos verbos típicos salvo "exhibir".

Según el proyectado artículo 173 bis CP, dicho contenido – ya sea audio, imagen o vídeo- debe ser generado, modificado o recreado mediante sistemas automatizados, *software*, algoritmos, inteligencia artificial o cualquier otra tecnología, de modo que parezca real. En este caso, debido al continuo avance y desarrollo de las tecnologías, el prelegislador ha decidido introducir una cláusula abierta que permite castigar como delito a quién difunda, exhiba o ceda un contenido hiperrealista manipulado mediante cualquier herramienta digital en el que se represente una conducta sexual o gravemente vejatoria.

Tal y como describe la figura delictiva proyectada, esos contenidos pueden ser generados, modificados o recreados mediante un algoritmo -entre otros medios-. Según la RAE, este término se define como "*un conjunto ordenado y finito de operaciones que permiten hallar la solución de un problema"* (Real Academia Española, s.f.). También se puede modificar la imagen a través de técnicas de inteligencia artificial, definida por ese mismo diccionario como "*disciplina que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico"* (Real Academia Española, s.f.).

Precisamente, el Reglamento (UE) 2024/1689, del Parlamento Europeo y del Consejo, en su artículo 3.1 establece que la inteligencia artificial es *"un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos*

[35] El artículo 197.7 CP dispone 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".*

explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales".

De igual modo, otra exigencia del tipo delictivo es que esos objetos materiales aparenten ser reales y, además, simulen una actuación de carácter sexual o gravemente vejatoria. Esto nos recuerda al concepto de pornografía infantil virtual regulada en el artículo 189 CP[36], en el que se establece que quedará subsumido en el tipo delictivo, *"todo material que represente de manera visual a un menor o una persona con discapacidad necesitada de especial protección participando en una conducta sexualmente explícita, real o simulada".*

La mención en el artículo proyectado "de modo que parezca real" nos conduce y como afirma Almenar Pineda (2025:123-124), a la Circular 2/2015, de la Fiscalía General del Estado donde se entiende por imágenes realistas aquellas que son "*cercanas a la realidad, a la que tratan de imitar. Dicho de otro modo, serían imágenes que no son reales pero lo parecen. Podrían abarcar imágenes alteradas de personas existentes e incluso imágenes generadas mediante ordenadores".* Por lo tanto, no se aplicaría en *"dibujos animados, manga o representaciones similares, pies no serían propiamente "imágenes realistas" en tanto no perseguirían ese acercamiento a la realidad".* De igual modo, los materiales simulados o manipulados han de representar situaciones de contenido sexual o gravemente vejatorias. Según la RAE, en relación con la primera de ellas, sería aquellas representaciones relacionadas con el sexo, con el *"conjunto de condiciones anatómicas y fisiológicas que caracterizan cada sexo"* (Real Academia Española, s.f.).

Respecto a las actuaciones "gravemente vejatorias", se trata de un concepto pendiente de valoración sobre el que la anterior Circular 2/2015 en un caso de utilización de menores para la elaboración de material pornográfico hace alusión a la STS 803/2010, de 30 de septiembre. Dicha resolución judicial introdujo criterios para interpretar este subtipo, subrayando su similitud con la agravación del antiguo artículo 180.1.1 CP -actual

[36] En el supuesto de que, finalmente, se aprobara el proyectado artículo 173 bis CP y las imágenes manipuladas correspondieran a menores de edad, se produciría un concurso aparente de normas penales con el delito de pornografía infantil virtual regulada en el artículo 189 CP.

artículo 180.1.2-[37] para las agresiones sexuales. Particularmente, a través de ella el Ato Tribunal argumentaba que *"el trato degradante o vejatorio equivale a realizar acciones con el fin de humillar, deshonrar, hacer despreciar o envilecer a alguien afectado a su dignidad humana".* En esa línea, la STS 344/2019, de 4 de julio de 2019, califica como gravemente vejatorias aquellas que poseen "*capacidad para humillar y rebajar a la víctima en el caso concreto"* (Almenar Pineda, 2025:124). Por ello, a modo de ejemplo, podría subsumirse o considerarse como imágenes vejatorias, las simulaciones de palizas contra mujeres (Lloria García, 2025:90), o las representaciones simuladas de actos sexuales entre mujeres y animales, así como aquellas en las que se muestran en actitudes vejatorias como puede ser pasearlas con correas como si fueran perros, enjaularlas como si fueran animales en un zoo o exhibirlas en escaparates como meros objetos en venta.

Asimismo, las imágenes o audios manipulados o generados mediante las TIC han de difundirse, revelarse o exhibirse sin la autorización de la persona afectada, es decir, aquella a la que se le representa mediante esos contenidos. En este sentido, coincido con Almenar Pineda (2025:124) en que la autorización únicamente será válida si la realiza la persona agraviada, en donde el consentimiento ha de ser bien de manera expresa o tácita, pero indubitada, tal y como confirma el criterio jurisprudencial para los delitos contra la intimidad. Por lo tanto, si un individuo otorga consentimiento para la difusión, cesión o exhibición de dichos materiales simulados mediante técnicas de IA generativa, la actuación no sería punible.

Por otro lado, en el plano de la culpabilidad, respecto al elemento subjetivo del tipo, el proyectado artículo 173 bis CP argumenta que, como elemento esencial para la consumación, ha de producirse *"un ánimo de menoscabar la integridad moral"*. Por ello, se requiere un dolo específico -a diferencia de lo que ocurre con los artículos 173.1 y 175 CP en los que simplemente se necesita un dolo genérico- (Almenar Pineda, 2025:125). De tal modo para poder aplicar el tipo delictivo proyectado se exige un propósito/una

[37] El artículo 180.1.2 CP dispone que:

> "*1. Las anteriores conductas serán castigadas, respectivamente, con las penas de prisión de dos a ocho años para las agresiones del artículo 178.1, de prisión de cinco a diez años para las agresiones del artículo 178.3, de prisión de siete a quince años para las agresiones del artículo 179.1 y de prisión de doce a quince años para las del artículo 179.2, cuando concurra alguna de las siguientes circunstancias: 2.ª Cuando la agresión sexual vaya precedida o acompañada de una violencia de extrema gravedad o de actos que revistan un carácter particularmente degradante o vejatorio".*

intencionalidad concreta, no es necesario únicamente que el sujeto activo difunda sin consentimiento unos contenidos sexuales o gravemente vejatorios hiperrealistas simulados, sino que se requiere que dicha difusión se haya llevado a cabo con ánimo de menoscabar la integridad moral de una/s persona/s.

5. ÍTER CRIMINIS

Respecto al *íter criminis,* la figura del proyectado artículo 173 bis CP no tipifica la conspiración, proposición y provocación para perpetrar el delito, por consiguiente, conforme los artículos 17.3 CP y 18.2 CP no se pueden castigar (Almenar Pineda, 2025:130).

Además, este tipo delictivo podría tratarse de un delito de resultado[38]. Por lo tanto, si consideramos dicho planteamiento, el delito no se consuma únicamente con la difusión, exhibición o cesión de una imagen o un audio de voz manipulado en el que se simule una situación sexual o gravemente vejatoria, se requiere la producción de un resultado (existiendo una separación espacio-temporal característica de esta clase de ilícitos penales) que, en este caso, menoscabe la integridad moral del sujeto pasivo (Almenar Pineda, 2025:130).

La naturaleza del delito de resultado conduce, necesariamente, a la posibilidad de castigar la tentativa. En relación con la tentativa acabada o inacabada[39] en aras a la imposición de la pena inferior en uno o dos grados a la indicada en el tipo delictivo proyectado y conforme al artículo 62 CP, expone Almenar Pineda (2025:130) fundamentándose en la STS 29 de octubre de 2021, que se produce tentativa acabada cuando "*los actos realizados, los instrumentos utilizados y la mecánica comisiva hubieran podido producir objetivamente el resultado perseguido según el plan del autor*". En cambio, el mismo autor expone que se ocasionaría tentativa inacabada en el

[38] Según Mir Puig (2016:230) para determinar si un delito es de resultado o de mera actividad se ha de atender si el tipo requiere o no que la acción vaya acompañada de la causación de un resultado que sea separable espacio-temporalmente de la conducta. En el supuesto de los delitos de mera actividad no resulta necesario. Sin embargo, en los delitos de resultado sí. (Mir Puig, 2016:230).

[39] A juicio de Diez-Ripollés (2020:552 y 560), la tentativa acabada presenta tres elementos: 1) la plena ejecución del delito; 2) la resolución delictiva o dolo; y 3) la no producción del resultado a pesar de que el autor no haya impedido de forma voluntaria su ocurrencia. En definitiva, en esta modalidad de tentativa se ha de haber ejecutado de modo completo el comportamiento típico peligroso. En cambio, la tentativa inacabada ostenta tres elementos: 1) el principio de ejecución del delito; 2) la resolución delictiva o dolo; y 3) la ausencia del resultado por interrupción voluntaria de la ejecución. (Díez-Ripollés, 2020:552-560).

supuesto de que el peligro para la integridad moral no se origine dado que tras generarse, modificarse o recrearse tecnológicamente el contenido sexual o gravemente vejatorio de la persona agraviada, el sujeto no sea capaz de difundirla, cederla o exhibirla. Situación distinta ocurre con la tentativa acabada cuando una vez difundida, exhibida o cedida no se ha logrado poner en riesgo el bien jurídico protegido. Asimismo, para tal propósito, se habrán tenido en cuenta los actos, los medios y el procedimiento utilizado por el individuo para efectuar la difusión, exhibición o cesión, así como la puesta en peligro de la integridad moral, por lo que, si con esos elementos hubiera sido posible producir el resultado de causación de peligro pretendido pero que no llega a producirse por causas ajenas a la voluntad del autor, podremos castigar por tentativa acabada. Además, la ausencia de puesta en peligro, por voluntad del autor, para el bien jurídico de la persona agraviada puede ocasionarse porque el sujeto activo haya determinado interrumpir de manera voluntaria la ultrafalsificación, produciéndose de esta manera el desestimiento o arrepentimiento activo en el caso de que el sujeto lleve a cabo la ejecución completa sin éxito voluntario[40] (Almenar Pineda, 2025:130-131).

Por otro lado, respecto a la tentativa inidónea[41], el contenido manipulado digitalmente ha de ser idóneo para poner en peligro la integridad moral del sujeto pasivo. Por ello esto ocasiona que en aquellos casos en los que los medios empleados no sean idóneos para la producción de ese riesgo, por ejemplo, si se pretende difundir, exhibir o ceder ese material en un chat cerrado en el que únicamente se encuentra el propio autor de los hechos (Almenar Pineda, 2025:131-132) o el contenido se lo envía a si mismo a través de una aplicación para móviles estaríamos ante un caso de tentativa inidónea del delito.

[40] El artículo 16.2 CP dispone que *"quedará exento de responsabilidad penal por el delito intentado quien evite voluntariamente la consumación del delito, bien desistiendo de la ejecución ya iniciada, bien impidiendo la producción del resultado, sin perjuicio de la responsabilidad en que pudiera haber incurrido por los actos ejecutados, si éstos fueren ya constitutivos de otro delito".* De tal modo y trasladándolo al estudio del proyectado artículo 173 bis CP, si una persona decide cometer un delito de difusión no consentida de un contenido *deepfake* sexual o gravemente vejatorio, pero finalmente de manera voluntaria decide detenerse *"bien desistiendo de la ejecución ya iniciada"* o *"bien impidiendo la producción de resultado"* -sin perjuicio de la responsabilidad que pudieran haber ocasionado los actos realizados- no será responsable penalmente del delito intentado, es decir, del ilícito penal proyectado.

[41] Diez-Ripollés (2020:563) expone que se produce tentativa inidónea cuando se ocasiona un comportamiento típico que no es peligroso. Entre los supuestos tradicionales de inidoneidad se hallan: 1) la inidoneidad en los medios; 2) la falta del objeto material del delito; y 3) la inidoneidad del sujeto activo del ilícito penal. (Díez-Ripollés, J.L., *Derecho Penal Español. Parte General*, 5ªed, Editorial Tirant Lo Blanch, 2020, p.552 y 5609).

6. AUTORÍA Y PARTICIPACIÓN

A juicio de Almenar Pineda (2025: 127-129), el delito del artículo 173 bis CP puede cometerse mediante los diversos tipos de autoría y participación, lo que implica tanto la autoría directa, la conjunta, la mediata como la inducción y la cooperación necesaria conforme al artículo 28 CP, así como a través de la complicidad en virtud del artículo 29 CP.

En primer lugar, el autor directo sería aquella persona que, con ánimo de menoscabar la integridad moral, difunde, exhibe o cede una imagen corporal o un audio de voz generado, modificado o recreado mediante sistemas automatizados, *software*, algoritmos, IA o cualquier otro tipo de tecnología que sea hiperrealista y simule acciones de contenido sexual o gravemente vejatorias.

En ese sentido, el tipo delictivo puede cometerse mediante coautoría, cuando los verbos típicos se cometan de manera concertada entre diversos sujetos, cada uno de ellos debe ostentar el dominio del hecho y participar como autores en la ejecución del delito. De igual manera, se puede perpetrar la conducta delictiva mediante autoría mediata, en aquellos casos en los que el dominio de la acción se lleva a cabo empleando violencia o intimidación sobre un individuo a efectos de que este cometa la conducta típica, o cuando el sujeto se halla en una situación de inimputabilidad ya sea por una anomalía o una alteración psíquica y otro individuo -el autor mediato- se aproveche de ello haciendo que este actúe como un instrumento del delito, o mediante la inducción a error a otro sujeto o incluso aprovechándose de la situación de error en la que se halla. Un ejemplo de error ocurriría si un individuo elabora un contenido simulado en el ordenador de otra persona con la finalidad de menoscabar la integridad moral de alguien y por ello, guarda el archivo en su memoria con un nombre lícito, y por error la persona titular de ese ordenador envía ese contenido a otro individuo pensando que aludía al título del archivo (Almenar Pineda, 2025:127-128).

Por otro lado, respecto a la posibilidad de cometer el delito mediante inducción es necesario analizar varios aspectos. En primer lugar, para que sea punible habría que demostrar, a través de los medios de prueba, que el sujeto que ha inducido a otro a llevar a cabo las conductas de difundir, exhibir o ceder una imagen o audio manipulado lo ha hecho con la finalidad de menoscabar la integridad moral de otro individuo. En este caso,

como ocurre en todos los delitos cometidos mediante inducción, debe constatarse que el inducido no pensaba previamente en cometer el delito. En segundo lugar, es necesaria la ejecución efectiva de la conducta típica ya que, de no ser así, se castigaría por proposición conforme el artículo 17.2 CP y al no expresarse explícitamente en el delito proyectado objeto de análisis sería atípica (Almenar Pineda, 2025:128). Por ejemplo, se aplicaría cuando un individuo a efectos de menoscabar la integridad moral de otra persona induce a otro sujeto -que previamente no había pensado- la idea de generar, modificar o recrear un contenido *deepfake* sexual o gravemente vejatorio a través de los medios descritos en el tipo delictivo proyectado y posteriormente dicho contenido lo difunda, ceda o exhiba.

De igual modo, el delito que se pretende regular en el artículo 173 bis CP se puede cometer mediante cooperación necesaria. Puede darse, por ejemplo, cuando una persona genera, modifica o recrea, mediante herramientas tecnológicas, una imagen o un audio de voz en la que se simula una conducta sexual o gravemente vejatoria, y posteriormente lo envía a un tercero para que este lo difunda entre otros individuos (Almenar Pineda, 2025:129). Otro ejemplo podría darse si una persona le presta su cuenta de una red social a otro sujeto para que este último a través de la cuenta prestada difunda, exhiba o ceda sin consentimiento una ultrafalsificación sexual o gravemente vejatoria.

Finalmente, la complicidad también encuentra cabida en la figura delictiva proyectada. Por ejemplo podría producirse en aquellos casos en los que un individuo de manera accesoria ayudara al sujeto activo en la generación, recreación o modificación de esos contenidos manipulados mediante técnicas de IA generativa, prestándole un dispositivo digital en el que se haya una aplicación o programa informático que posibilita llevar a cabo dicha acción, aunque el autor directo sería quién *a posteriori* genera, recrea o modifica y quien posteriormente difunde, exhibiera o cediera esos materiales simulados (Almenar Pineda, 2025:129).

7. PENAS

De acuerdo con el artículo 173 bis CP que pretende introducir el Proyecto de Ley, se castigará con la pena de prisión de uno a dos años a quién, con ánimo de atentar contra la integridad moral de una persona, difunda, ceda o exhiba una imagen corporal o audio de voz generado, modificado o recreado por sistemas automatizados, *software*,

algoritmos, inteligencia artificial o cualquier otra tecnología, en los que se simule una situación sexual o gravemente vejatoria.

En ese sentido, el proyectado tipo delictivo prevé una agravación de la pena de 1 y 6 meses a 2 años de prisión cuando los sujetos pasivos sean menores de edad o personas con discapacidad necesitada de especial protección[42].

Por otro lado, respecto al tipo agravado, la figura delictiva proyectada del artículo 173 bis CP dispone que "*se aplicará la pena en su mitad superior si dicho material ultrafalsificado se difunde a través de un medio de comunicación social, por medio de internet o mediante el uso de tecnologías, de modo que aquel se hiciera accesible a un elevado número de personas en el espacio virtual*"[43].

Finalmente, el párrafo III del artículo 173 bis CP regula que "*Las autoridades judiciales ordenarán la adopción de las medidas necesarias para la retirada de los contenidos a los que se refieren los párrafos anteriores, para la interrupción de los servicios que ofrezcan predominantemente dichos contenidos o para el bloqueo de unos y otros cuando radiquen en el extranjero".*

Del mismo modo, desde un punto de vista procesal-penal, el párrafo III del proyectado artículo es parecido al artículo 13 LECrim en el que se dispone que "*En la instrucción de delitos cometidos a través de internet, del teléfono o de cualquier otra tecnología de la información o de la comunicación, el juzgado podrá acordar, como primeras diligencias, de oficio o a instancia de parte, las medidas cautelares consistentes en la retirada provisional de contenidos ilícitos, en la interrupción provisional de los servicios que ofrezcan dichos contenidos o en el bloqueo provisional de unos y otros cuando radiquen en el extranjero".*

Además, también resulta similar al párrafo II de los delitos de promoción, fomento o incitación, a través de las TIC al suicidio (artículo 143 bis CP), a las autolesiones

[42] La redacción inicial que contenía el Anteproyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales no recogía la agravante de la pena cuando los sujetos pasivos son personas menores de edad o personas con discapacidad necesitadas de especial protección ni tampoco comprendía el párrafo tercero del proyectado artículo 173 bis. Por todo lo demás, dicha figura delictiva coincidía en su totalidad con la que recoge el Proyecto de Ley Orgánica.

[43] El citado tipo agravado del proyectado artículo 173 bis CP se analizará en profundidad a partir del epígrafe noveno del presente capítulo.

(artículo 156 ter CP), al consumo de sustancias nocivas para la salud (artículo 361 bis CP) a los menores de edad o personas con discapacidad necesitadas de especial protección o a cometer determinados delitos sexuales contra ellas (artículo 189 bis CP)[44].

A juicio de Muñoz Conde (2023:668), en relación con los anteriores delitos tecnológicos estas medidas pueden ser incluso más efectivas que las penas sobre todo si se pueden adoptar como medidas cautelares a lo largo del proceso penal. Precisamente, en ocasiones, debido a la complejidad que supone la identificación de los autores o la imposibilidad de someterlos a la jurisdicción penal española es posible que no se pueda ni siquiera imponer las condenas.

8. PROBLEMAS CONCURSALES

El proyectado artículo 173 bis CP presenta algunos problemas concursales con otros ilícitos penales. Sin embargo, dadas las características del presente trabajo simplemente me limitaré a anunciarlas y examinarlas en un sentido amplio.

A juicio de Almenar Pineda (2025:132-133), en primer lugar, si relacionamos la creación de *deepfakes* pornográficas o vejatorias con el tipo básico del artículo 173.1 CP, resulta necesario tener en cuenta diversos aspectos. En los casos en los que se produzca un trato degradante contra otra persona, afectando así a su integridad moral, el hecho delictivo se castigará mediante el artículo 173 CP. Por ejemplo, existen resoluciones judiciales emitidas por nuestro Tribunal Supremo[45] en las que se ha castigado a través del artículo 173.1 CP casos de grabación y difusión de imágenes gravemente vejatorias en las que se cosificaban a menores de edad o personas especialmente vulnerables. Sin embargo, cuando se trate de un contenido de carácter sexual o gravemente vejatorio que haya sido manipulado o generado por medio de sistemas informáticos con el ánimo de menoscabar la integridad moral de otro individuo, sería aplicable el artículo 173 bis CP

[44] El párrafo II del artículo 143 bis CP relativo a la promoción, fomento o incitación al suicidio a menores de edad o personas con discapacidad necesitadas de especial protección señala que *"Las autoridades judiciales ordenarán la adopción de las medidas necesarias para la retirada de los contenidos a los que se refiere el párrafo anterior, para la interrupción de los servicios que ofrezcan predominantemente dichos contenidos o para el bloqueo de unos y otros cuando radiquen en el extranjero".* Los tres delitos restantes conllevan también esta misma previsión en sus respectivos párrafos segundos.

[45] Entre las más destacadas se halla la STS 547/2022, de 2 de junio de 2022, en la que un individuo le dio a una persona galletas rellenas de pasta dentrífica y posteriormente la grabó mientras las ingería para después, con ánimo lucrativo, subirla a su canal de *Youtube*.

en virtud del principio de prevalencia de la ley penal especial sobre la común o genérica, conforme el artículo 8.1 CP (Almenar Pineda, 2025:132-133).

Otro problema concursal que destaca Almenar Pineda (2025: 133-134) es la relación entre el proyectado delito del artículo 173 bis CP relativo a la difusión sin consentimiento de *deepfakes* sexuales o vejatorias del artículo 173 bis CP y el delito del artículo 172 ter 5 CP que castiga "*el que, sin consentimiento de su titular, utilice la imagen de una persona para realizar anuncios o abrir perfiles falsos en redes sociales, páginas de contacto o cualquier medio de difusión pública, ocasionándole a la misma situación de acoso, hostigamiento o humillación".* El artículo 173 bis CP se vincula con el 172 ter 5 CP en relación con la ausencia de consentimiento y la difusión de imagen ajena. No obstante, se diferencian en los medios y el objeto, ya que en el proyectado artículo la imagen ha de ser corporal y simulada, modificada o generada a partir de sistemas informáticos, simulando hechos de contenido sexual o gravemente vejatorios y no es requisito la concurrencia de una situación de hostigamiento, acoso o humillación sino que requiere un ánimo de atentar contra la integridad moral de un individuo. De igual modo, el bien jurídico también es diverso, a través del artículo 172 ter 5 CP se atenta contra la libertad y en el proyectado artículo 173 bis CP contra la integridad moral. Por todo ello, subsumir la conducta en un tipo delictivo o en otro dependerá necesariamente del objeto material, de los medios y del bien jurídico protegido que el sujeto activo intente menoscabar.

Por otro lado, resulta relevante incidir en las diferencias entre el delito de *sexting* (197.7 CP) y el delito proyectado del artículo 173 bis CP. El primero de ellos, tal y como se ha expuesto previamente, se castiga a quién sin consentimiento de la persona agraviada, difunda, revele o ceda a terceros imágenes o grabaciones audiovisuales que hayan sido obtenidas con consentimiento en un domicilio o en cualquier otro lugar fuera del alcance de la mirada de terceros pero su difusión -no consentida- menoscaba gravemente la intimidad personal del sujeto pasivo. En este caso, esos materiales representan a una imagen real de una persona y por ello, se trata de un delito contra la intimidad. Sin embargo, en el proyectado artículo 173 bis CP, los contenidos son simulados. Por lo tanto, por ejemplo, en el supuesto de que se difunda sin consentimiento una imagen sexual íntima que represente un cuerpo real de un individuo se aplicaría el delito *sexting* pero si el contenido es generado, modificado o recreado a través de técnicas

de IA generativa, la conducta se subsumiría en la figura delictiva proyectada en el artículo 173 bis CP. Todo ello ocasiona que, para diferenciar un tipo delictivo de otro, atendamos al bien jurídico protegido y al objeto material del delito.

Por otro lado, pueden surgir problemas concursales entre el delito del artículo 173 bis CP y el delito de injurias. La Exposición de Motivos del Proyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales argumenta que los delitos de injurias se subsumen en el tipo delictivo a través del principio de consunción. En particular, dispone que *"pues ha de tomarse en cuenta no solo la afectación a la autoestima y la heteroestima, sino también la cosificación e instrumentalización que se produce sobre el sujeto pasivo, generalmente mujeres y niñas, niños y adolescentes que son tratados como objetos de consumo."*. En esta línea la conducta se castigará a través del proyectado artículo 173 bis CP si concurren los elementos recogidos en el mismo. En cambio, falta alguno de ellos se podría castigar el hecho como un delito de injurias regulado en el artículo 208 CP (Almenar Pineda: 2025, 134-135).

En último lugar, un supuesto significativo es que en el caso de que se difunda una ultrafalsificación sexual en la que se representen a menores de edad, se produce un concurso de normas penales dado que el artículo 189 CP regula la pornografía virtual infantil. Por ello, en un futuro resultará necesario distinguir qué diferencias -si es que las hay- existen entre un término y otro.

9. EL TIPO AGRAVADO DEL PROYECTADO ARTÍCULO 173 BIS

El tipo agravado del proyectado de artículo 173 bis CP en su párrafo II argumenta que "*Se aplicará la pena en su mitad superior si dicho material ultrafalsificado se difunde a través de un medio de comunicación social, por medio de internet o mediante el uso de tecnologías, de modo que aquel se hiciera accesible a un elevado número de personas en el espacio virtual*".

Concretamente el prelegislador fundamenta su regulación en el apartado IX de la Exposición de Motivos del Proyecto de Ley Orgánica, disponiendo: *"Además de que las ultrafalsificaciones generalmente se difunden en el ciberespacio, con la potencialidad de*

permanencia que ello implica, como se ha advertido respecto de los delitos tecnológicos de contenido, se produce un aumento de la lesividad en relación con otras modalidades de ataque por la enorme dificultad de distinguir entre el contenido falso y el real debido a la precisión de las nuevas tecnologías y por el mayor grado de veracidad que mantenemos respecto de materiales audiovisuales sobre materiales escritos".

Siguiendo a Almenar Pineda (2025:135), respecto al fundamento del tipo agravado, coincido con que la agravante del artículo proyectado resulta parecida a la que dispone el artículo 578.2 CP[46] para los delitos de terrorismo, a las regulada en el artículo 599.2[47] para el descubrimiento y revelación de secretos e informaciones relativas a la Defensa Nacional y, en sentido amplio, a la agravante en los delitos de calumnias e injurias en los que se castiga con mayor pena cuando se cometa la conducta típica con publicidad conforme al artículo 211 CP. A juicio de García Rivas (2016:586), ello es así debido a que cuanto mayor sea la difusión de dichos contenidos mayor será la lesión al honor.

Del análisis realizado se desprende la idea de que el prelegislador opta por aumentar la penalidad en aquellos casos en los que los contenidos manipulados se difundan. En esa línea, el Tribunal Supremo a través de diversas resoluciones judiciales ha dejado de manifiesto su preocupación sobre la evolución tecnológica y como esta en ocasiones puede aumentar la lesividad sobre los bienes jurídicos protegidos, requiriendo así que una respuesta del sistema penal. Precisamente, la STS 106/2015, de 19 de febrero de 2015 argumenta que *"cualquier política criminal hoy en día no puede ignorar esta explosión tecnológica que permite divulgar cualquier mensaje en pocos segundos a una multitud de usuarios situados en países lejanos con los que se obtiene una publicidad de los mensajes impensable hace años"* (Devís Matamoros, 2022:2).

De ese modo la STS 4/2017, de 18 de enero de 2017 dispone que *"la extensión actual de las tecnologías al servicio de la comunicación intensifica de forma exponencial del daño de afirmaciones o mensajes que, en otro momento, podrían haber limitado sus*

[46] El artículo 578.2 CP dispone que *"Las penas previstas en el apartado anterior se impondrán en su mitad superior cuando los hechos se hubieran llevado a cabo mediante la difusión de servicios o contenidos accesibles al público a través de medios de comunicación, internet, o por medio de servicios de comunicaciones electrónicas o mediante el uso de tecnologías de la información".*

[47] El artículo 599.2 CP argumenta que "*La pena establecida en el artículo anterior se aplicará en su mitad superior cuando concurra alguna de las circunstancias siguientes: 2) Que la revelación consistiera en dar publicidad al secreto o información en algún medio de comunicación social o de forma que asegure su difusión".*

perniciosos efectos a un reducido y seleccionado gripo de destinatarios. Quien hoy incita a la violencia en una red social sabe que su mensaje se incorpora en las redes telemáticas con vocación de perpetuidad. Además, carece de control sobre su zigzagueante difusión, pues desde que ese mensaje llega a manos de su destinatario éste puede multiplicar su impacto mediante sucesivos y renovados actos de transmisión" (Almenar Pineda, 2025:136).

Siguiendo a Lloria García (2020:94-95), las características de los medios tecnológicos ocasionan un incremento del injusto respecto de los ataques que derivan del ámbito analógico, por diversas razones: por su anonimato, su viralidad y su permanencia (estas dos últimas también son señaladas en la ya citada STS 4/2017, de 18 de enero de 2017). En esa línea, "la viralidad" es decir, la difusión de la información a través de la red, conlleva un incremento cuantitativo en la lesión -que puede incluso una transformación cualitativa de la misma- dado que ocasiona que cualquier acto alcance a un conjunto indeterminado -prácticamente ilimitado- de personas en poco tiempo. Respecto a "la permanencia" (o "vocación de perpetuidad"), dada la naturaleza propia de internet y de las propias redes sociales resulta complejo que un contenido previamente difundido desaparezca completamente, provocando de esta manera que persista la lesión.

Acerca de la conducta que dispone el tipo agravado del proyectado artículo 173 bis CP resulta conveniente reflexionar sobre varios aspectos. En primer lugar, se aplicará a todos los casos que recoge el tipo básico regulado en el párrafo primero, dado que el tipo delictivo argumenta que *"si dicho material ultrafalsificado se difunde a través de".* De ello, tal como plantea Almenar Pineda (2025:135) el tipo básico también hace alusión a quienes difundan un contenido *deepfake* sexual o gravemente vejatorio. Por lo tanto, resulta necesario distinguir, por un lado, los supuestos de difusión del tipo básico y por otro, los de tipo agravado. Se ha analizado que difundir un contenido implica comunicar algo a otras personas. A ello, el tipo agravado concreta que se aplicará la pena en su mitad superior si dicha conducta se lleva a cabo "*a través de un medio de comunicación social, por medio de internet o mediante el uso de tecnologías, de modo que aquel se hiciera accesible a un elevado número de personas en el espacio virtual"*. De tal modo, en aras de determinar qué tipo delictivo resulta de aplicación habrá que atender a los medios utilizados. Por ello, si para ejecutar el delito se utilizan los instrumentos que describe párrafo segundo, se castigaría a través del tipo agravado. En cambio, si no se

usan, la conducta se subsumiría en el tipo básico conforme al párrafo I (Almenar Pineda, 2020:135).

Por lo tanto, se producirá un aumento de la pena cuando los contenidos se difundan a través de un medio de comunicación social, Internet o el uso de tecnologías. Respecto al término "Internet", según la RAE se entiende como una *"Red informática mundial, descentralizada, formada por la conexión directa entre computadoras mediante un protocolo especial de comunicación"* (Academia de la Lengua Española, s.f.). Acerca del concepto "comunicación social", el informe "Medios de Comunicación Social" elaborado por el Ministerio de Sanidad y Consumo (1994:15 y 18) define que *"es aquel que permite difundir el mensaje dirigido al conjunto de la sociedad o a una parte de la misma"* y además, existen diversas clases, en particular pueden ser medios impresos o medios audiovisuales. En esa línea, Almenar Pineda (2025:137) hace referencia a la STS 1074/2004, de 18 de octubre de 2004, en la que se deduce que en oposición al resto de medios (uso de tecnologías e Internet) los medios de comunicación social "*cumplen con una necesaria labor profesional, amparada en su derecho fundamental a difundir información (esencial en un Estado social y democrático de derecho".*

En último lugar, el prelegislador, tal y como lleva a cabo en la redacción del tipo básico, a través del concepto "uso de tecnologías" introduce una cláusula abierta en la que el medio tecnológico en cuestión haga accesible el contenido *deepfake* "*a un elevado número de personas en el espacio virtual*". En esta línea, se aplicará el tipo agravado del artículo proyectado cuando se utilicen los medios concretos anteriormente descritos y en consecuencia, el contenido hiperrealista sexual o gravemente vejatorio manipulado pueda ser visualizado por una gran cantidad de personas en el mundo virtual (Almenar Pineda, 2025: 137-138).

Por todo ello, del análisis del tipo agravado del proyectado artículo 173 bis CP se concluye que, desde un punto de vista jurídico-penal, resulta adecuada la decisión del prelegislador de agravar la pena cuando los materiales simulados se difunden a través de un medio de comunicación social, Internet o el uso de tecnologías, dado que, tanto por la vocación de permanencia como por la viralidad que caracteriza a los delitos tecnológicos así como, por el potencial alcance a un conjunto indeterminado que implica la acción "difundir", se produce un incremento de la lesividad que exige de un mayor reproche penal.

CONCLUSIONES

PRIMERA.- Según su interpretación gramatical, la inteligencia artificial es una disciplina científica que permite crear programas informáticos capaces de llevar a cabo actividades comparables a las que realizan los humanos, como el aprendizaje, el razonamiento y la percepción. Por consiguiente, dadas sus características se ha convertido en un instrumento idóneo para atentar o poner en peligro bienes jurídicos necesitados de protección penal. Por ello, en el ámbito comunitario se ha aprobado, entre otros textos, el Reglamento (UE) 2024/1689, del Parlamento y del Consejo, de 13 de junio, por el que se establecen normas armonizadas en materia de inteligencia artificial.

SEGUNDA.- Las *deepfakes,* también denominadas ultrafalsificaciones o ultrasuplantaciones, son contenidos audiovisuales falsos hiperrealistas, elaborados a través de técnicas de *deep learning*. Debido a sus características y funcionamiento, un uso ilícito de este tipo de materiales es susceptible de vulnerar derechos fundamentales, ya que pueden emplearse para la difusión de desinformación, alterar procesos electorales, manipular pruebas en procedimientos judiciales, o como medios comisivos para la perpetración de delitos normalmente de carácter económico. Además, pueden ser utilizados para la creación y difusión sin consentimiento de contenidos sexuales y/o vejatorios, especialmente destinados contra mujeres, niñas y niños.

TERCERA.- En el ámbito comunitario, conscientes del creciente avance y del peligro que ocasiona la difusión de *deepfakes* sexuales y/o vejatorias, se ha aprobado la Directiva 2024/1385, del Parlamento y del Consejo, de 14 de mayo de 2024, sobre la lucha contra la violencia contra las mujeres y contra la violencia doméstica, a través de ella se insta a los Estados miembros a que regulen este tipo de conductas delictivas pero únicamente haciendo referencia a los contenidos manipulados de carácter sexual.

A su vez, en el ámbito nacional, se ha aprobado el Proyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales, en el que, entre otras medidas, desea incorporar el artículo 173 bis CP a efectos de tipificar y castigar dichas actuaciones como delitos contra la integridad moral.

CUARTA.- Las ultrafalsificaciones afectan desproporcionadamente a las mujeres debido a que pueden ser empleadas como un mecanismo de control y dominación, al despojarlas del uso de su propio cuerpo y tratarlas como un mero objeto consumible. De tal modo resulta razonable entender que la difusión no consentida de *deepfakes* sexuales y/ o vejatorias se ha convertido en una nueva manifestación de violencia de género digital.

QUINTA.- Generalmente cuando las víctimas de este tipo de contenidos son mujeres con relevancia política y/o social o celebridades estas actuaciones se pueden vincular con la manosfera o machosfera, comunidades de hombres que a través de Internet difunden y comparten contenido misógino de mujeres, normalmente con relevancia política o social puesto que ven en peligro su hegemonía masculina. Por otro lado, cuando las víctimas son mujeres anónimas, las *deepfakes* pueden utilizarse como una forma de chantaje a efectos de cometer sextorsión o "pornografía de la venganza".

SEXTA.- A través de la redacción del proyectado artículo 173 bis CP se desprende la idea de que este tipo de contenidos simulados sexuales o vejatorios afectan a la integridad moral del sujeto que se representa en dichos materiales.

Por lo tanto, mediante su tipificación no se protege el bien jurídico "intimidad" ya que para atentar o poner en peligro dicho valor debe tratarse de cuerpos reales, situación que no ocurre con las ultrafalsficaciones en las que se representan cuerpos generados, modificados o recreados digitalmente. En esa línea, tampoco se atenta o se pone en peligro el bien jurídico "honor" en exclusividad, ya que si fuera así no se castigarían otras acciones, como aquellas, que tienen propósitos económicos o meramente degradantes. En consecuencia, a mi juicio resulta acertada la postura del prelegislador al entender que a través del principio de consunción se subsume el honor en el bien jurídico "integridad moral" por considerarse este último más amplio.

SÉPTIMA.- En el hipotético caso de que el proyectado artículo 173 bis CP finalmente se tipificara y en el supuesto de que las víctimas de contenidos *deepfakes* sexuales fueran menores de edad se produciría un concurso de normas penales ya que el artículo 189 CP castiga el delito de pornografía infantil virtual. Por lo tanto, a efectos de evitar confusiones jurídico-penales, sería conveniente esclarecer que diferencias (si es

que las hay) ostenta este concepto y las ultrafalsificaciones de carácter sexual cuando el titular del material sea una persona menor de edad.

A su vez, habría que plantearse también la cuestión relativa al tratamiento que debe otorgarse a la imagen del cuerpo de la actriz porno utilizada para elaborar la ultrafalsificación. La principal reflexión consiste en determinar si dicha persona sería también sujeto pasivo del delito proyectado o si, por el contrario, la conducta únicamente conllevaría en una afectación del derecho de propiedad intelectual.

OCTAVA.- El fenómeno de los *deepfakes* es uno de los desafíos más complejos y urgentes en la sociedad digital actual, tanto por sus características como por los riesgos jurídicos que ocasiona. Por lo tanto, resulta necesario llevar a cabo un análisis en profundidad de estas acciones, así como de sus consecuencias con el fin de establecer la posibilidad de sanción en aquellos supuestos que supongan una vulneración de bienes jurídicos dignos, susceptibles y necesitados de la protección del Derecho Penal.

BIBLIOGRAFÍA

1. MONOGRAFIAS

Almenar Pineda, F. (2018). *Ciberdelincuencia.* Editorial Juruá.

Díez-Ripollés, J.L. (2020). *Derecho Penal español. Parte General* (5ª. ed.). Tirant lo Blanch.

Hernández Peña, J.C. (2022). *El marco jurídico de la Inteligencia Artificial. Bases para su diseño: principios, técnicas, procedimientos y estructuras de gobernanza* (1ª. ed.). Aranzadi.

Lloria García, P. (2020). *Violencia sobre la mujer en el siglo XXI. Violencia de control y nuevas tecnologías. Habitualidad, sexting y stalking* (1ª. ed). Iustel.

Mir Puig, S. (2016). *Derecho Penal. Parte General* (10ª. ed.). Reppertor.

Muñoz Conde, F. (2023). *Derecho Penal. Parte Especial* (25.ª ed.). Tirant Lo Blanch.

Rodríguez Mourullo, G. (1978). *Derecho Penal. Parte General*. Civitas.

Russell, S.J y Norvig, P. (2008). *Inteligencia Artificial. Un enfoque moderno* (2.ª ed.). Pearson.

Varona, G. (2025). *Victimología didáctica y aplicada: análisis de casos.* Laborum Ediciones.

Vestri, G. (dir.)., Castilla Barea, M (Coord.)., Castellano Claramunt, J., García Rodríguez, M., Goenechea Permisán, C., Herce Maza, J.I., Montes, C., Rebato Moreno, M., Mota Sánchez, E.M., Pérez Merina, M., Rodíl Garrido, A., Sánchez Sánchez., E.M & Wisner Glusko, C (2023). *Diccionario de términos para comprender la transición digital* (1.ª ed.). Aranzadi.

2. CAPÍTULOS DE LIBROS EN OBRAS COLECTIVAS

Ayllón García, J.D. (2022). La Inteligencia Artificial como medio de difusión y control de las *fakes news* (1º. ed.). En C. Villegas Delgado (Coord.) *El derecho en la encrucijada tecnológica: Estudios sobre derechos fundamentales, nuevas tecnologías e inteligencia artificial* (218-238). Tirant Lo Blanch.

Doval Pais, A. (2016). Delitos contra la intimidad, el derecho a la propia imagen y a la inviolabilidad del domicilio (1). Delitos contra la intimidad y los datos personales. En J. Boix Reig (Dir.), *Derecho Penal. Parte Especial Volumen I (La protección penal de los intereses jurídicos personales. (Adaptados a la reforma de 2015 del Código Penal*) (493-550). Iustel.

García Rivas, N. (2016). *Delitos contra el honor.* En J. Boix Reig (Dir.). *Derecho Penal. Parte Especial Volumen I (la protección penal de los intereses jurídicos personales. (Adaptados a la reforma de 2015 del Código Penal)* (567-606). Iustel.

Grima Lizandra, V. (2016). Delitos contra la integridad moral. En J. Boix Reig (Dir.) *Derecho Penal. Parte Especial. Volumen I (La protección penal de los intereses jurídicos personales (Adaptados a la reforma de 2015 del Código Penal)* (291-328). Iustel.

Lloria García, P. (2023). Cibercriminalidad en tiempos de confinamiento. En G. González Agudelo & L.R. Ruíz Rodríguez (Dirs.). *La política criminal de excepción en Iberoamérica durante la crisis pandémica*. (191-200). Tirant Lo Blanch.

Lloria García, P. (2024). Violencia digital de género ¿algo más que una afectación a la intimidad? El caso Buturugă vs Rumania. En J. Llobet Rodríguez & L.R. Ruíz Rodríguez. *Derechos Humanos y tecnologías en la jurisprudencia de la Corte IDH y del TEDH* (173-200). Editorial Jurídica Continental.

3. ARTÍCULOS CIENTÍFICOS

Aguayo, L., & Ruiz del Olmo, F.J. (2024). Vídeos falsos y desinformación ante la IA: el deepfake como vehículo de la posverdad. *Revista de Ciencias de la Comunicación e Información,* 29, 1-14. https://doi.org/10.35742/rcci.2024.29.e294

Alfonso Sánchez, I.R. (2016). La Sociedad de la Información, Sociedad de Conocimiento y Sociedad del Aprendizaje. Referentes en torno a su formación. *Bibliotecas. Anales de Investigación*, 12 (2), 235-243.

Almenar Pineda, F. (2025). Las ultrafalsificaciones como delito en el anteproyecto de ley orgánica para la protección de las personas menores de edad en los entornos digitales. *Revista Ius Criminale*, 2, 85-144. https://doi.org/10.69592/3045-6681-N2-ABRIL

Arana, C. (2021). Redes neuronales recurrentes: Análisis de los modelos especializados en datos secuenciales, Serie Documentos de Trabajo, No. 797. *Universidad del Centro de Estudios Macroeconómicos de Argentina.*

Barba Arteaga, C. (2024). *Deepfakes* sexuales: impacto, prevención y perspectivas de género en el entorno digital, 15(2), 229-244. *Mhjournal (Miguel Hernández Communication Journal).* https://doi.org/10.21134/zt4eht31

Barrientos-Báez, A., & Piñeiro Otero, T., & Porto Renó, D., (2024). Imágenes falsas, efectos reales. Deepfakes como manifestaciones de la violencia política de género. *Revista Latina de Comunicación Social*, 82, 01-29. https://www.doi.org/10.4185/RLCS-2024-2278

Beada, I. (2016). Acerca del concepto de *dignidad humana* en la filosofía kantiana: del hombre como *fin en sí mismo* al hombre como *ciudadano del mundo*. *Revista Estudios Kantianos*, 1,24-38.

Blázquez, R. (2023). Deepfakes en el procedimiento probatorio. *Revista Vasca de derecho procesal y arbitraje*, 35 (3), 223-256.

Bustamante Donás, J. (2001) Hacia la cuarta generación de Derechos Humanos: repensando la condición humana en la sociedad tecnológica. *Revista Iberoamericana de Ciencia, Tecnología, Sociedad e Innovación*, 1, 1-18.

Citron, D., & Chesney B. (2018). Deep Fakes: A Looming Challenge of Privacy, Democracy and National Security. *Scholarly Commons at Boston University School of Law,* 107, 1753-1820. https://doi.org/10.15779/Z38RV0D15J

De la Cuesta Aguado, P. (2024). El Derecho Penal frente al Reglamento de Inteligencia Artificial. *Revista Ius Criminale*, 1, 83-107. https://doi.org/10.69592/3045-6681-N1-OCTUBRE-2024-ART-4

Devís Matamoros, A. (2022). Aproximación jurisprudencial sobre la difusión de contenidos delictivos en redes sociales. *Diario La Ley*, (101153).

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

Dolz Lago, M.J. (2022). Una aproximación jurídica a la Inteligencia Artificial en *Revista Diario La ley,* (10096).

Farenc Lendvai, G., & Gosztonyi, G. (2024). *Deepfake* y desinformación -¿Qué puede hacer el derecho frente a las noticias falsas creadas por *deepfake*? *Revista de Internet, Derecho y Política,* 41*,* 1-13. 10.7238/idp.v0i41.427515

García-Mingo, E., & Díaz Fernández, S., & Tomás-Forte, S. (2022). (Re)configurando el imaginario sobre la violencia sexual desde el antifeminismo: el trabajo ideológico de la manosfera española. *Política y Sociedad*, 59(1), 1-15. https://dx.doi.org/10.5209/poso.80369

García Ull, F.J. (2021). Deepfakes: el próximo reto en la detección de noticias falsas. *Anàlisi: Quaderns de Comunicació i Cultura*, 64, 103-120. https://doi.org/10.5565/rev/analisi.3378

Gil Cruz, E.M. (2024). Principales cuestiones del Reglamento de Inteligencia Artificial (IA) de la UE. *Unión Europea Aranzadi*, 8-9.

Gil Lloria, M. (2025). Desinformación e inteligencia artificial en procesos electorales. *El Cronista del Estado Social y Democrático de Derecho*, 116, 80-92

Ging, D., & Siapera, E., (2018). Special issue on online misogyny. *Feminist Media Studies*, 18 (4), 515-524. 10.1080/14680777.2018.1447345

González Pulido, I. (2023). El uso de la inteligencia artificial generativa en la investigación de la ciberdelincuencia de género: ante el auge de los *deepfakes*. *Revista Ius Et Scientia*, 9 (2), 157-180. https://dx.doi.org/10.12795/IETSCIENTIA

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

Jacobsen, B.N., y Simpson, J. (2024). The tensions of deepfakes. *Information, Communication and Society*, 27, 6, 1095-1109. https://doi.org/10.1080/1369118X.2023.2234980

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

Lacalle, C., & Gómez Morales, B & Vicent-Ibáñez, M., (2023). Misoginia y construcción de masculinidades tóxicas en la Manosfera española *(Burbuja.info). Profesional de la información*, 32(2). *https://doi.org/10.3145/epi.2023.mar.15*

Lloria García, P. (2025). Ciberdelincuencia de género. Especial referencia a las ultrafalsificaciones o deepfakes. *Cuadernos Penales José María Lidón*, 19, 69-94.

Maddocks, S., (2020). A Deepfake Porn Plot Intended to Silence Me: exploring continuities between pornographic and "political" deepfakes. *Porn Studies,* 7(4), 415-423. https://doi.org/10.1080/23268743.2020.1757499

Mantilla, K., (2013). Gendertrolling: Misogyny Adapts to New Media. *Feminist Studies*, 39 (2), 563-570.

McCarthy J. (2007). What is artificial intelligence?. *Standford University*. https://www-formal.stanford.edu/jmc/whatisai.pdf

McGlynn, C., & Tuna Toparlak., (2025). The "new voyeurism": criminalizing the creation of "deepfake porn". *Journal of Law and Society*, 52, 204-228. DOI: 10.1111/jols.12527

Mihailova M. (2021). To Dally with Dalí: Deepfake (Inter)faces in the Art Museum.*The International Journal of Research into New Media Technologies*, 27(4), 882-898. https://doi.org/10.1177/13548565211029401

Montesinos García, A., (2024). Inteligencia Artificial en justicia con perspectiva de género: amenazas y oportunidades. *Revista Actualidad Jurídica Iberoamericana*, 21, 566-597.

Newton B.O., & Stanfill M. (2020). My NSFW video has partial occlusion: deepfakes and the technological production of non- consensual pornography. *Porn Studies*, 7(4), 398-414. 10.1080/23268743.2019.1675091

Nilsson N.J. (2010). The quest for artificial intelligence. A history of ideas and achievements. Cambridge University Press. https://ai.stanford.edu/~nilsson/QAI/qai.pdf

Öhman., C. (2020). Introducing the pervert´s dilemma: a contribution to the critique of Deepfake Pornography. *Ethics and Information Technology*, 22, 133-140. https://doi.org/10.1007/s10676-019-09522-1

Okolie, C. (2023). Artificial Intelligence-Altered Videos (*Deepfakes*), Image-Based Sexual Abuse, and Data Privacy Concerns. *Journal of International Women's Studie*s, 25(2), 1-16.

O´Shea, K., & Nash, R. (2015). An introduction to Convolutional Neural Networks. *Revista Arvix,* 1-11. https://doi.org/10.48550/arXiv.1511.08458

Pérez-Úgena, M. (2024). La Inteligencia Artificial: definición, regulación y riesgos para los derechos fundamentales. *Estudios de Deusto. Revista de Derecho Público,* 72(1), 307-337. https://doi.org/10.18543/ed.3108

Rodríguez Andrés, R. (2018). Fundamentos del concepto de desinformación como práctica manipuladora en la comunicación política y las relaciones internacionales. *Historia y comunicación social,* 23 (1), 231-244. https://doi.org/10.5209/HICS.59843

Santana Soriano, E., & Báez Vizcaíno, K. (2022). Ciberspacio y cibermundo: delimitaciones conceptuales desde el materialismo sistémico. *Ciencia y Sociedad*, 47(1), 45-57. https://doi.org/10.22206/cys.2022.v47i1.pp45-57

Santiesteban Galarza, M. (2024). La criminalización de as ultrafalsificaciones (con especial atención a las implicaciones de la normativa europea de servicios digitales e inteligencia artificial). *Revista de Derecho Penal y Criminologia*, 3.ª Época, 31, 211-249. https://doi.org/10.5944/rdpc.ENERO.2024.38890

Serrano Acitores, A. (2023). Consideraciones éticas y legales en el uso de la Inteligencia Artificial: el enfoque de la Unión Europea. *Revista de fiscalidad internacional y negocios transnacionales* 24, 69-106.

Simó Soler, E. (2023). Retos jurídicos derivados de la Inteligencia Artificial Generativa. Deepfakes y violencia contra las mujeres como supuesto de hecho. *Revista InDret*, 2, 493-515. 10.31009/InDret.2023.i2.11

Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social media + society.* 1-13. https://doi.org/10.1177/2056305120903408

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

Villar-Aguilés, A., & Pecourt Gracia, J., (2021). Antifeminismo y troleo de género en Twitter. Estudio de la subcultura trol a través de #STOPfeminazis. *Revista de Cultura Digital y Movimientos Sociales*, 18 (1), 33-44. http://dx.doi.org/10.5209/TEKN.70225

Villarino Marzo, J. (2017). Cuarta generación de derechos: reflexiones sobre la libertad de expresión en Internet. *Revista de las Cortes Generales*, (100-102), 47-99. https://doi.org/10.33426/rcg/2017/100-102/27

4. ARTÍCULOS DE PRENSA

Aguayo, D. (2024). Encuentran a más de 200 idols femeninas del K-pop en sitios ilegales de deepfakes. *La Cuarta Diario Pop.* https://www.lacuarta.com/mundo/noticia/encuentran-a-mas-de-200-idols-femeninas-del-k-pop-en-sitios-ilegales-de-deepfake/D6MNRVA4P5BZNB3Q2QS7SJFS5Q/

Allyn, B. (2022). Deepfakes video of Zelensky could be "tip of the iceberg" in info war, experts warn. *Diario NPR.* https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

Ayyub, R. (2018). I was the victim of a deepfake porn plot intended to silence me. *Huffpost*.https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316

De Tena Piera, C.L. (2024). Los procesos electorales de 2024, un riesgo por los "deepfakes". *Diario Ethic*. https://ethic.es/2024/02/los-procesos-electorales-de-2024-en-riesgo-por-los-deepfakes/

Díaz, J. (2021). Un Tom Cruise digital indistinguible del real se hace viral en TikTok. *El Confidencial.* https://www.elconfidencial.com/tecnologia/novaceno/2021-02-26/tom-cruise-tiktok-deep-fake_2969284/

Gómez Isaza, DR. (2025). La historia de Hannah Grundy, la australiana víctima de deepfake por un amigo: "Convirtió cada uno de mis recuerdos en porno". *El colombiano.com.* https://www.elcolombiano.com/internacional/la-historia-de-hannah-grundy-victima-de-deepfake-con-inteligencia-artificial-IG26570109

Kong, H. (2024). Everyone looked real': multinational firm's Hong Kong office loses HK$200 million after scammers stage deepfake video meeting. *South China Morning Post* https://www.scmp.com/news/hong-kong/law-and-crime/article/3250851/everyone-looked-real-multinational-firms-hong-kong-office-loses-hk200-million-after-scammers-stage

La Vanguardia (2024). Estafa con "deepfakes": 24 millones de euros y una reunión virtual con una sola persona real. *La Vanguardia.*https://www.lavanguardia.com/tecnologia/ciberseguridad/20240205/9513300/estafa-multiples-deepfakes-24-millones-euros-persona-real.html

Martin, T.W. (2024). 'Anyone Can Be a Victim': Sprawling AI Fake Nudes Crisis Hits South Korea. *The Wall Street Journal.* https://www.wsj.com/world/asia/anyone-can-be-a-victim-sprawling-ai-fake-nudes-crisis-hits-south-korea-eec02232

Merino, M. (2025). "Publicas tus momentos más felices y alguien los convierte en porno" Hannah descubrió que un "amigo" creaba deepfakes con sus fotos. *Genbeta.* https://www.genbeta.com/inteligencia-artificial/publicas-tus-momentos-felices-alguien-convierte-porno-hannah-descubrio-que-amigo-creaba-deepfakes-sus-fotos

Sebastian, M. (2024). AI and deepfakes blur reality in India elections. *BBC*. https://www.bbc.com/news/world-asia-india-68918330

Turnbull T., (2025). “Cada momento publicado en Instagram o Facebook lo convirtió en porno”. Cómo un hombre traicionó a su amiga usando su imagen con inteligencia artificial. *BBC News Mundo*. https://www.bbc.com/mundo/articles/cjdegvm50g4o

UN News (2022). Halt all retaliation attacks against Indian journalist Rana Ayyub – UN experts. *UN News*. https://news.un.org/en/story/2022/02/1112362

West, K. (2024). “Mi mejor amigo colgó fotos pornográficas falsas con mi cara en internet”. *BBC News Mundo*. Disponible en: https://www.bbc.com/mundo/articles/cl4kxe0xkk6o

5. ARTÍCULOS EN REVISTAS DIGITALES

Cole, S. (2017). AI-Assisted Fake Porn Is Here and We´re All Fucked. *Vice*. https://www.vice.com/en/article/gal-gadot-fake-ai-porn/

O´Neill. L. (2024). Fotos falsas, consecuencias reales: Alexandria Ocasio-Cortez y la batalla contra el porno IA. *Rolling Stone.* https://es.rollingstone.com/fotos-falsas-consecuencias-reales-alexandria-ocasio-cortez-y-la-batalla-contra-el-porno-por-ia/

6. INFORMES

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). *The estate of deepfakes: landscape, threats, and impact.* Deeptrace. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

EUROPOL (2020). *Malicious Uses and Abuses of Artificial Intelligence.* https://www.europol.europa.eu/publications-events/publications/malicious-uses-and-abuses-of-artificial-intelligence

EUROPOL (2024). *Facing reality? Law enforcement and the challenge of deepfakes.* https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes

Home Security Heroes (2023). *2023 State of deepfakes.* https://www.securityhero.io/state-of-deepfakes/

Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial (2019). *Directrices éticas para una IA fiable.* https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

Instituto Europeo de la Igualdad de Género (2017). *La ciberviolencia contra las mujeres y niñas,* en.https://eige.europa.eu/sites/default/files/documents/ti_pubpdf_mh0417543esn_pdfweb_20171026164000.pdf

Ministerio de Igualdad (2022). *La violencia política por razón de género en España: una aproximación cualitativa.* https://violenciagenero.igualdad.gob.es/violenciaencifras/estudios/investigaciones/violencia-politica/

Ministerio de Igualdad (2025). *Violencia contra las mujeres, niñas, niños y adolescentes en el ámbito digital.* https://www.igualdad.gob.es/comunicacion/sala-de-prensa/el-ministerio-de-igualdad-hace-publico-el-estudio-violencia-contra-mujeres-ninas-ninos-y-adolescentes-en-el-ambito-digital/

Ministerio de Sanidad y Consumo (1994). *Medios de comunicación social.* https://www.dsca.gob.es/sites/default/files/53_mediosComunicacion.pdf

Observatorio Nacional de Tecnología y Sociedad (2022). *Políticas públicas contra la violencia digital de género 2022.* Ministerio de Asuntos Económicos y Transformación Digital. https://www.ontsi.es/sites/ontsi/files/2022-04/violenciadigitalgenerounarealidadinvisible_2022.pdf

Observatorio Nacional de Tecnología y Sociedad (2023). *El uso de las tecnologías por los menores en España. Edición 2024- datos 2023.* Observatorio Nacional de Tecnología y Sociedad. Red.es. Secretaría de Estado de Digitalización e Inteligencia Artificial. Ministerio para la Transformación Digital y de la Función Pública. https://www.ontsi.es/sites/ontsi/files/2024-07/El%20uso%20de%20menores%20en%20Espa%C3%B1a%202023.pdf

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

UNESCO (2023). *Kit de herramientas global sobre IA y el estado de derecho para el poder judicial.* https://unesdoc.unesco.org/ark:/48223/pf0000387331_spa

7. DICCIONARIOS

Diccionario Panhispánico del español jurídico. *Ciberespacio.* En Dpej.rae.es. Recuperado el 20 de mayo de 2025, en https://dpej.rae.es/lema/ciberespacio

Merriam-Webster (s.f.). *Deepfake.* En Merriam-Webster.com Dictionary. Recuperado el 18 de febrero de 2025, en https://www.merriam-webster.com/dictionary/deepfake

Oxford English Dictionary (s.f.). *Disinformation*. En Oed.com. Recuperado el 20 de mayo 2025. https://www.oed.com/search/dictionary/?scope=Entries&q=disinformation

Real Academia de Lengua Española (s.f.). *Ciberespacio*. En RAE.com Diccionario. Recuperado el 19 de mayo de 2025, en https://dle.rae.es/ciberespacio

Real Academia de Lengua Española (s.f.) *Inteligencia Artificial*. En RAE.com Diccionario. Recuperado el 20 de marzo de 2025, en https://dle.rae.es/inteligencia

Real Academia de Lengua Española (s.f.). *Sexualidad.* En RAE.com Diccionario. Recuperado el 4 de julio de 2025, en https://dle.rae.es/sexualidad

Seon (2025). *¿Qué son las deepfakes?* En Glosario. Recuperado el 20 de marzo de 2025, en https://seon.io/es/recursos/glosario/deepfake/

OTROS DOCUMENTOS

Agencia Española de Protección de Datos (2023). *La violencia contra las mujeres y niñas aglutina el 70% de los casos que se denuncian en el Canal prioritario.* Disponible en: https://www.aepd.es/prensa-y-comunicacion/notas-de-prensa/violencia-digital-contra-mujeres-y-ninas-aglutina-70-casos-canal-prioritario

Fiscalía General del Estado (2015). Circular 2/2015, de 19 de junio, sobre los delitos de pornografía infantil tras la reforma operada por Ley Orgánica 1/2015, *Doctrina de la Fiscalía General del Estado.* FIS-C-2015-00002

Naciones Unidas (UNRIC) (2024). *¿Puede la inteligencia artificial influenciar los procesos electorales?*https://unric.org/es/peligros-y-beneficios-de-la-inteligencia-artificial-en-procesos-electorales/

BLOGS O MEDIOS ESPECIALIZADOS/CORPORATIVOS/ TÉCNICOS

Anikin, A. (2023). *Deepfakes de voz: no creas todo lo que escuchas*. Kaspersky daily. https://www.kaspersky.es/blog/audio-deepfake-technology/29025/?srsltid=AfmBOootYvqrSbs9oulxqOdBbeRJwZell6tZtksuYfh580hiLizJANrO

Aware Biometrics (2024). *Cómo la biometría ayuda a defenderse contra los deepfakes en el aprendizaje y las certificaciones en línea.* Consultado 21 de marzo 2025. https://www.aware.com/es/como-la-biometria-ayuda-a-defenderse-contra-los-deepfakes-en-el-aprendizaje-y-las-certificaciones-en-linea-blog/

Beamonte, P. (2018). *FakeApp, el programa de moda para crear vídeos porno falsos con IA*. Hipersexual. https://hipertextual.com/2018/01/fakeapp-videos-porno-falsos-ia

Cole, S. (2024). *Laws About Deepfakes Can´t Leave Sex Workers Behind.* 404 Media. https://www.404media.co/laws-about-deepfakes-cant-leave-sex-workers-behind/

EDJ Tech Law School (2023). *Deepfakes: manipulación de pruebas y su impacto en los procesos judiciales.* https://www.edjxtechlawschool.com/post/deepfakes-manipulaci%C3%B3n-de-pruebas-y-su-impacto-en-los-procesos-judiciales

Foro Económico Mundial (2022). *¿Cómo combatir el preocupante aumento del uso de las deepfakes en la ciberdelincuencia?* https://es.weforum.org/stories/2023/05/como-combatir-el-preocupante-aumento-del-uso-de-deepfakes-en-la-ciberdelincuencia/

Micucci, M. (2024). *Herramientas para detectar deepfakes y combatir la desinformación.* Welivesecurity. https://www.welivesecurity.com/es/seguridad-digital/herramientas-para-detectar-deepfakes-combatir-desinformacion/

Nebrada, P (2023). *Qué es un deepfake y cómo pararlo gracias a la biometría*. https://alicebiometrics.com/deepfake-que-es-como-pararlo-biometria/

Salako, E (2023). *Herramientas de inteligencia artificial para detectar deepfakes.* International Journalists´Network. https://ijnet.org/es/story/herramientas-de-inteligencia-artificial-para-detectar-deepfakes

JURISPRUDENCIA

Sentencia del Tribunal Supremo 1725/2001 (sala de lo penal) de 3 de octubre de 2001

Sentencia del Tribunal Supremo 1074/2004 (sala de lo penal) de 18 de octubre de 2004

Sentencia del Tribunal Supremo 233/2009 (sala de lo penal, sección 1º) de 3 de marzo de 2009 (Recurso 1449/2008)

Sentencia del Tribunal Supremo 803/2010 (sala de lo penal, sección 1º) de 30 de septiembre de 2010 (Recurso 644/2010)

Sentencia del Tribunal Supremo 325/2013 (sala de lo penal, sección 1º) de 2 de abril de 2013 (Recurso 1148/2012)

Sentencia del Tribunal Supremo 106/2015 (sala de lo penal) de 19 de febrero de 2015 (Recurso 1374/2014)

Sentencia del Tribunal Supremo 1619/2016 (sale de lo penal) de 18 de enero de 2017 (Recurso 1619/2016)

Sentencia del Tribunal Supremo 344/2014 (sala de lo penal, sección 1º) de 4 de julio de 2019 (Recurso 396/2019)

Sentencia del Tribunal Constitucional 426/2021 (sala segunda) de 19 de mayo de 2021.

Sentencia del Tribunal Supremo 829/2021 (sala de lo penal) de 29 de octubre de 2021 (Recurso 4684/2019)

Sentencia del Tribunal Supremo 547/2022 (sala segunda, sección pleno) de 2 de junio de 2022 (recurso 547/2022)

NORMATIVA

a. Ámbito internacional

Convenio de Estambul. Consejo de Europa, 11 de mayo (*Council of Europe Convention on preventing and combating violence against women and domestic violence*). Ratificado por España el 18 de mayo de 2014, publicado en BOE el 6 de junio de 2014, entrando en vigor de forma general y para España el 1 de agosto de 2014. https://rm.coe.int/1680462543

b. Ámbito comunitario

Comisión Europea. (2018a). *Inteligencia Artificial para Europa*. Comunicación COM (2018) 237 final, de 25 de abril de 2018.

Comisión Europea (2018b). *Plan coordinado sobre inteligencia artificial.* Comunicación COM (2018) 795 final, de 7 de diciembre de 2018. https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=COM:2018:795:FIN

Comisión Europea (2019). *Generar confianza en la inteligencia artificial centrada en el ser humano.* COM (2019) 168 final, de 8 de abril de 2019. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52019DC0168&from=PT

Comisario de Derechos Humanos del Consejo de Europa (2019). *Unboxing de la Inteligencia Artificial: 10 pasos para proteger los derechos humanos*, de 14 de mayo de 2019. https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64

Comisión Europea (2020). *Libro Blanco sobre la Inteligencia Artificial: un enfoque europeo orientado a la excelencia y la confianza.* Comunicación COM (2020) 65 final, de 19 de febrero de 2020. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065

Comisión Europea (2021). *Fomentar un enfoque europeo de la inteligencia artificial. Comunicación* COM (2021) 205 final, de 21 de abril de 2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021DC0205

Directiva (UE) 2024/1385 del Parlamento Europeo y del Consejo, de 14 de mayo de 2024, sobre la lucha contra la violencia contra las mujeres y la violencia doméstica. *Diario Oficial de la Unión Europea*. «DOUE», 1385, de 24 de mayo de 2024, páginas 1-36. DOUE-L-2024-80770. https://www.boe.es/buscar/doc.php?id=DOUE-L-2024-80770

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). *Diario Oficial de la Unión Europea*, «DOUE» L. 1689, de 12 de julio de 2024, páginas

1-144. DOUE-L-2024-81079. https://www.boe.es/buscar/doc.php?id=DOUE-L-2024-81079

Consejo Europeo y Consejo de la Unión Europea. Reglamento de IA. Consultado el 21 de marzo de 2025. https://www.consilium.europa.eu/es/policies/artificial-intelligence/

c. Ámbito nacional

Real Decreto de 14 de septiembre de 1882 por el que se aprueba la Ley de Enjuiciamiento Criminal, «BOE» núm. 260, de 17 de septiembre de 1882. https://www.boe.es/eli/es/rd/1882/09/14/(1)/con

Constitución Española. «BOE» núm 311, de 29 de diciembre de 1978. https://www.boe.es/eli/es/c/1978/12/27/(1)/con

Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal, «BOE» núm. 281, de 24 de noviembre de 1995. https://www.boe.es/eli/es/lo/1995/11/23/10/con

Ley Orgánica 8/2021, de 4 de junio, de protección integral a la infancia y la adolescencia frente a la violencia., «BOE» núm. 134, de 5 de junio de 2021. https://www.boe.es/eli/es/lo/2021/06/04/8/con

Anteproyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales. *Ministerio de Presidencia, Justicia y relaciones con las cortes*

Proyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales, *Boletín Oficial de las Cortes Generales*, núm 52-1, de 11 de abril de 2025. https://www.congreso.es/public_oficiales/L15/CONG/BOCG/A/BOCG-15-A-52-1.PDF