LA CRIMINALIZACIÓN DE LAS ULTRAFALSIFICACIONES (CON ESPECIAL ATENCIÓN A LAS IMPLICACIONES DE LA NORMATIVA EUROPEA DE SERVICIOS DIGITALES E INTELIGENCIA ARTIFICIAL)

Mario Santisteban Galarza

Investigador predoctoral en la Universidad del País Vasco (UPV/EHU)[1]

Title: *The criminalisation of deep fakes (with particular attention to the implications of the European regulation on Digital Services and Artificial Intelligence)*

Resumen: La Inteligencia Artificial presenta riesgos que están siendo atendidos por distintas ramas del ordenamiento jurídico. Una subespecie de estos sistemas, los llamados modelos generativos, presentan una particularidad como es que pueden crear todo tipo de contenidos, entre ellos ultrafalsificaciones (comúnmente conocidos como *deep fakes*), esto es, representaciones de personas realizando comportamientos que no tuvieron lugar en un principio. Recientemente la criminalización de las ultrafalsificaciones se ha planteado tanto en el ordenamiento jurídico comunitario como el nacional. El presente trabajo analiza la respuesta que las citadas reformas presentan ante las ultrafalsificaciones de carácter sexual, enmarcándolas en el complejo marco regulatorio europeo de las

[1] La publicación de este trabajo es parte del proyecto Ius-Machina, sobre las bases normativas y el impacto real de la utilización de algoritmos predictivos en los ámbitos judicial y penitenciario TED2021-129356B-I00, financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea «NextGenerationEU»/PRTR y del proyecto de I+D+i PID2021-125730OBI00, acrónimo #FakePenal, financiado por MCIN/AEI/10.13039/501100011033/ y por «FEDER Una manera de hacer Europa». Asimismo, es consecuencia de la Convocatoria PIB/2020 de la Universidad del País Vasco (UPV/EHU).

ultrafalsificaciones, que establece obligaciones adicionales a los operadores de estos sistemas relevantes para el Derecho penal[2].

Palabras clave: Ultrafalsificaciones; Inteligencia Artificial; IA generativa; Reglamento de servicios digitales; Reglamento de IA.

Abstract: *Artificial Intelligence presents risks that are being addressed by different enforcement techniques. A subspecies of these systems, the generative models, enable the creation of all kinds of content, including deep fakes, representations of people performing behaviors that did not take place in the first place. Recently, the criminalization of deep fakes has been raised in both EU and national legal systems. This paper analyzes the response that the aforementioned reforms present to sexual deep fakes, framing them in the complex European regulatory framework of deep fakes, that establishes additional obligations to the operators of these systems relevant to criminal law.*

Keywords: *Deep Fakes; Artificial Intelligence; generative AI; Digital Services Act; AI regulation.*

1. Introducción

La promesa de un mayor rol de la Inteligencia Artificial[3] (IA), y en general de la automatización algorítmica en nuestra sociedad, se ve como una oportunidad para encarar ciertos problemas como la seguridad, la lucha contra el cambio climático o la mejora de la medicina, pero también como un riesgo para el modelo de convivencia basado en las liber-

[2] Me gustaría agradecer al profesor Fernando Miró Llinares por su consejo y apoyo en la realización de este artículo. Asimismo, me gustaría agradecerle la oportunidad de asistirle en la elaboración del borrador de informe general de la Asociación Internacional de Derecho Penal (AIDP), XXI Congreso Internacional de Derecho Penal «Inteligencia Artificial y Sistema de Justicia», Sección II (Parte Especial), proceso enormemente enriquecedor para mi trayectoria investigadora y que ha informado especialmente este trabajo.

[3] Sobre la discusión del concepto de Inteligencia Artificial en Derecho Penal L., Picotti, «Traditional Criminal law categories and AI: crisis or palingenesis? General report», *RIDP*, Vol. 94 issue 1, 2023, págs. 11-53.

tades constitucionales[4]. De ahí que exista una fuerte discusión donde frecuentemente podemos encontrar posiciones que exacerban tanto los aspectos positivos y negativos de estos avances[5], frente a una posición del legislador europeo aún en construcción.

Especial complejidad presentan los retos de la llamada Inteligencia Artificial Generativa. Esta se integra por modelos avanzados de aprendizaje automático que se entrenan para generar nuevos datos, como texto, imágenes o audio, lo que la diferencia de modelos predictivos o comparativos tradicionales[6]. Son modelos especializados en interactuar con los usuarios y crear contenidos en función de peticiones específicas. Dentro de estos sistemas podemos distinguir aquellos que permiten crear ultrafalsificaciones (conocidas por el termino anglosajón *deep fakes*), que podemos definir preliminarmente como contenido sintético que muestra a personas realizando conductas que no llegaron a realizar. Esta tecnología presenta beneficios en el ámbito educativo, y también como medio de expresión artística y satírica[7]. No obstante, su potencial disruptivo es innegable. Las ultrafalsificaciones suponen la amenaza de que contenidos y canales a través de los que nos comunicamos diariamente, ya sea audio, video o incluso texto, sean «subvertidos»[8], creando incertidumbre en una sociedad enormemente digitalizada. Más aún, son una amenaza contra bienes jurídicos de relevancia penal, teniendo en cuenta uno de sus usos más problemáticos: la creación de imágenes pornográficas de forma no consentida.

Por poner algunos ejemplos recientes, en España, personalidades como Rosalía[9] y Laura Escanes ya han sido víctimas de esta tecnología[10]. Asimismo, menores han corrido la misma suerte, creándose imágenes pornográficas de los mismos. El hecho de que fueran sus compañeros de

---

[4] Comisión Europea, «Libro blanco sobre la inteligencia artificial. Un enfoque europeo orientado a la excelencia y la confianza», 2020, Disponible en: https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065&from=ES (consultado 11. 11. 2023).

[5] F. Miró Llinares, «Inteligencia artificial, delito y control penal: nuevas reflexiones y algunas predicciones sobre su impacto en el derecho y la justicia penal», *El Cronista del Estado Social y Democrático de Derecho*, 100, 2022, págs. 174-183.

[6] P. Hacker, A. Engel y M. Mauer, «Regulating ChatGPT and other Large Generative AI Models», *ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*, June 12–15, 2023, p. 2.

[7] D. Citron y R. Chesney, «Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security», *California Law Review*, 197, 2019, p. 1769 y ss.

[8] H. Ajder, G. Patrin, F. Cavalli y L. Cullen, 'The State of Deepfakes: Landscape, Threats, and Impact», 2019 https://regmedia.co.uk/2019/10/08/deepfake_report.pdf (consultado 11. 11. 2023).

[9] https://www.eldiario.es/blog/micromachismos/rosalia-fotos-manipuladas-mostraban-desnuda-cuerpo-mujer-no-mercancia-estrategia-marketing_132_10235525.html (consultado 11.11.2023).

[10] https://www.vozpopuli.com/espana/deepfake-arrasa-internet-contenido-adultos-generado-ia-cada-vez-tiene-mas-presencia.html (consultado 11.11.2023).

clase los que «crearon» estas imágenes, contratando un servicio a través de Internet, alerta de la creciente expansión del acceso a esta tecnología y los riesgos que acarrea. También se reportan supuestos en los que se han utilizado para crear pornografía infantil simulada[11].

Atendiendo a esta preocupación, se ha planteado recientemente la criminalización de ciertos usos de las ultrafalsificaciones en diferentes iniciativas tanto a nivel nacional como comunitario. Como señala Lloria García, la tipificación de nuevas conductas al hilo del cambio tecnológico puede ser necesaria pues «no siempre los instrumentos tradicionales del derecho penal son válidos para resolver las cuestiones que surgen a propósito de las lesiones a bienes jurídicos nuevos o la afectación de los tradicionalmente tutelados con una mayor intensidad»[12]. Esta operación no es sencilla, existiendo en muchos casos dudas sobre los bienes jurídicos afectados por los comportamientos cometidos a través de las nuevas tecnologías[13], y diferentes instrumentos supranacionales que empujan a la criminalización que pueden crear disrupciones en el ordenamiento nacional[14].

Muy ligado a lo anterior, la criminalización de ciertos usos de las ultrafalsificaciones, específicamente las que tienen un carácter sexual, presenta el problema de la identificación del bien jurídico afectado por la creación y difusión del contenido simulado. Esta es una decisión de gran trascendencia, que no solo determinará la ubicación sistemática del tipo propuesto en el Código, sino que ayudará a dotar de un «contenido material» al injusto, sin el cual no puede apreciarse la tipicidad[15]. Por el contrario, en el estadio en el que se encuentra la reflexión dogmática sobre las ultrafalsificaciones sexuales difícilmente puede abordarse la problemática con rigor sin hacer un tratamiento monográfico del tema que desplace a otras cuestiones. Así, en esta primera aproximación se opta por no detenerse en este importante debate para abordar otros aspectos no menos relevantes en el proceso de criminalización. Y es que uno de los principales problemas a la hora de plantear la criminalización de ciertos usos de las ultrafalsificaciones en nuestro ordenamiento es la

[11] https://elpais.com/sociedad/2022-12-21/detenido-un-pederasta-que-usaba-inteligencia-artificial-para-crear-material-de-abuso-sexual-infantil.html (consultado 11.11.2023).

[12] P. Lloria García, «Delitos y redes sociales: los nuevos atentados a la intimidad, el honor y la integridad moral». Especial referencia al «sexting», *La Ley Penal*, Nº 105, Sección Estudios, Noviembre-Diciembre 2013, págs. 1-10.

[13] N. J. De La Mata Barranco, «Reflexiones sobre el bien jurídico a proteger en el delito de acceso informático ilícito (art. 197 bis cp)», *Cuadernos de política criminal*, Núm. 118, 2016, págs. 43-86.

[14] A. Galán Muñoz, «La internacionalización de la represión y la persecución de la criminalidad informática: un nuevo campo de batalla en la eterna guerra entre prevención y garantías penales», *Revista Penal*, Núm. 24, 2009, pp. 90-107.

[15] P. Sánchez-Ostiz, *A vueltas con la Parte Especial (Estudios de Derecho Penal)*, Atelier, Barcelona, 2020, p. 39 y ss.

existencia paralela de otros textos normativos, que fijan obligaciones a los operadores de ultrafalsificaciones, o que incluso plantean también la criminalización.

Teniendo esto en cuenta, el presente trabajo realiza una panorámica de las propuestas normativas más relevantes en la regulación de las ultrafalsificaciones, tratando de aportar reflexiones relevantes que guíen el inminente proceso de criminalización de algunos de sus usos. Para ello, el apartado segundo analiza brevemente el funcionamiento de la tecnología que sustenta las ultrafalsificaciones, así como los usos problemáticos más comunes identificados por la doctrina. El apartado tercero se ocupa de las dos propuestas de criminalización más relevantes, y el posicionamiento que adoptan frente a un tipo de ultrafalsificación: las que recrean imágenes de contenido sexual de forma no consentida. El apartado cuarto analiza la normativa extrapenal que fija deberes a distintos sujetos implicados en el ciclo de vida de las ultrafalsificaciones, discutiendo su relevancia a la hora de determinar la responsabilidad penal de los distintos operadores de las ultrafalsificaciones. Finalmente se esbozan unas breves conclusiones.

2. Funcionamiento y usos problemáticos de las ultrafalsificaciones

El término *deep fake*[16] aúna los dos términos ingleses *deep* (profundo) y *fake* (falso). En el argot jurídico regulatorio europeo se está utilizando el término «ultrafalsificación» en castellano para referirse a este tipo de contenidos, por el cual nos decantaremos en este trabajo. Las ultrafalsificaciones pueden definirse como contenido audiovisual manipulado o sintético en el que aparecen personas que parecen decir o hacer algo que nunca han dicho o hecho, producido mediante técnicas de inteligencia artificial, incluido el aprendizaje automático y el aprendizaje profundo[17]. Como veremos, la propuesta de Reglamento único de Inteligencia Artificial utiliza una definición muy similar[18]. No obstante,

[16] Su uso popular se remonta al año 2017, cuando un usuario de Reddit publicó vídeos pornográficos alterados digitalmente, en el que los rostros de las actrices se sustituían por el de celebridades como Taylor Swift, Scarlett Johansson y Gal Gadot. UK Centre For Data Ethics And Innovation, «Snapshot Paper - Deepfakesand Audiovisual Disinformation», 2020, Disponible en: https://www.gov.uk/government/publications/cdei-publishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai/snapshot-paper-deepfakes-and-audiovisual-disinformation, p. 4. (consultado 11.11.2023).

[17] Parlamento Europeo, «*Tackling deepfakes in European policy*», 2021, Disponible en: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2021)690039, p. 2.

[18] «Un contenido de sonido, imagen o vídeo manipulado o sintético que puede inducir erróneamente a pensar que es auténtico o verídico, y que muestra representaciones de personas que parecen decir o hacer cosas que no han dicho ni hecho, producido utilizando

otros autores introducen matices divergentes, por ejemplo, prefiriendo no referirse a tecnologías en específico y centrándose en el resultado: un material (imágenes o texto) creado con medios técnicos avanzados en el que una persona parece estar haciendo o diciendo algo que en realidad no hizo o dijo, y en el que la manipulación tiene una posibilidad considerable de burlar el juicio de una persona o un sistema de detección basado en la IA[19].

Las ultrafalsificaciones admiten diversas formas de manipulación sobre el contenido. Estas pueden basarse en: tomar una imagen del rostro de una persona y añadirla a la de otra persona (*face replacement*); la recreación del rostro, variando diferentes características sin implicar un «cambio de cara» (*fase reenactment*); la generación de rostros, que crea imágenes de caras que no se corresponden con ninguna imagen real (*face generation*); o el sintetizado de voz, donde se utiliza un software avanzado para crear un modelo de la voz de alguien (*speech synthesis*)[20]. Como puede entreverse los tipos de contenido generados son amplios, abarcando las imágenes, estáticas o en movimiento, pero también el audio[21]. A pesar de ello lo cierto es que las ultrafalsificaciones que están despertando especial interés regulatorio son aquellas referidas a las imágenes estáticas o en movimiento, a las cuáles les prestaremos especial atención aquí.

Las herramientas de edición de foto o video se han encontrado disponibles desde hace mucho tiempo; pensemos en el caso de Photoshop, que ha permitido manipular imágenes desde finales del siglo pasado. A nadie se le escapa que este tipo de ediciones manuales son costosas, requiriendo cierto tiempo y pericia para llegar a resultados satisfactorios[22]. Avances en distintos campos tecnológicos han permitido automatizar dichos procesos. Es el caso de algoritmos que permiten detectar automáticamente patrones en el rostro de las personas, dando lugar a la tecnología del reconocimiento facial[23]. También ha tenido un importante juego la aparición de Internet y servicios relacionados con la Web. 2.0. Los efectos «generativos» de Internet son conocidos, permitiendo un ciclo de

técnicas de IA, incluido el aprendizaje automático y el aprendizaje profundo» artículo 3 44 quinquies. La versión enmendada por el Parlamento Europeo puede encontrarse aquí: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html (consultado 11.11.2023).

[19] B. Van Der Sloot y Y. Wagensveld, «Deepfakes: regulatory challenges for the synthetic society», *Computer Law & Security Review*, Volume 46, 2022, p. 4.

[20] UK Centre For Data Ethics And Innovation, *op.cit*, p. 3.

[21] Incluso podemos llegar a hablar ultrafalsificaciones de texto, pero estas se obtienen a través del uso de algoritmos muy distintos que imitan el lenguaje natural Parlamento Europeo, *op.cit*, pág. 13.

[22] J. Ice, «Defamatory Political Deepfakes and the First Amendment», *Case W. Rsrv. L. Rev*, Vol. 417, 2019, Disponible en: https://scholarlycommons.law.case.edu/caselrev/vol70/iss2/12 (consultado 11.11.2023).

[23] Parlamento Europeo, *op,cit,* p. 7.

retroalimentación que facilita la innovación a través de la modificación del software, facultad a disposición de quien disponga de un ordenador personal[24]. Así, aplicaciones especializadas en generar ulrafalsificaciones son variantes de otros modelos generativos de imágenes de código abierto. Por otro lado, Internet, y Web 2.0 han permitido nutrir a las aplicaciones de un conjunto de imágenes de fácil acceso. Más aún, autores defienden que la facilidad en la difusión de la información en este medio y la supuesta existencia de burbujas informativas hacen del ciberespacio un entorno en el cual las ultrafalsificaciones tienen muchas posibilidades de prosperar[25].

Las ultrasificaciones más avanzadas beben de estos cambios tecnológicos, y se fundamentan en el aprendizaje automático (*machine learning*), algoritmos que permiten mejorar su rendimiento a través de la detección de patrones de un conjunto de datos y de las redes neuronales. Concretamente, de un tipo de red neuronal que se denomina red generativa adversativa (*generative adversarial networks*). En esta coinciden dos sistemas, uno denominado generativo y otro llamado discriminador. El generativo crea un contenido tratando de que este se asemeje lo más posible a los datos de entrada. El contenido es después analizado por el sistema discriminador, que tiene la misión de determinar si este proviene de la base de datos o del modelo generativo[26]. Se establece un ciclo de retraolimentación y el sistema generativo se nutre de las decisiones del discriminador, detectando nuevos patrones que pueden afinar su rendimiento. Nótese, no obstante, que otros sistemas pueden estar involucrados, por ejemplo, aplicando varios procesos a los datos de entrada para facilitar la tarea del generador y el discriminador[27]. Asimismo, pueden utilizarse otras técnicas menos complejas, comúnmente denominadas *shallow face* y no dependientes del aprendizaje automático[28].

Como hemos apuntado en la introducción, las ultrafalsificaciones pueden ser utilizadas para realizar comportamientos problemáticos, en algunos casos aptos para lesionar bienes jurídicos relevantes para el Derecho Penal. Sin ánimo de exhaustividad, y en la línea de otros estudios, podemos identificar tres usos problemáticos de las ultrafalsificación: su

---

[24] En este sentido el influyente artículo de J. Zittrain, «The Generative Internet», *Harvard Law Review*, Vol. 119:1974, 2006.

[25] D. Citron y R. Chesney, *op.cit*, p. 1766 y ss.

[26] Europol Innovation Lab, «Facing reality? Law enforcement and the challenge of deepfakes», 2022, disponible en: https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes p. 8. (consultado 11.11.2023).

[27] Para una explicación amena del proceso véase https://www.youtube.com/watch?v=aMlcd8SFF-U (consultado 11.11.2023).

[28] V. Ciancaglini, C. Gibso y D. Sancho, «Malicious Uses and Abuses of Artificial Intelligence», Trend Micro Research, 2020, https://www.europol.europa.eu/cms/sites/default/files/documents/malicious_uses_and_abuses_of_artificial_intelligence_europol.pdf p. 54 y ss. (consultado 11. 11. 2023).

uso con propósitos desinformativos, para cometer delitos contra el patrimonio y la creación de imágenes sexuales de forma no consentida.

Empezando con el primero, la desinformación es considerada en el marco de la Unión Europea como «información verificablemente falsa o engañosa que, se crea, presenta y divulga con fines lucrativos o para engañar deliberadamente a la población y puede causar un perjuicio público, entendido como amenazas contra los procesos democráticos políticos y de elaboración de políticas, así como contra los bienes públicos, como la protección de la salud, el medio ambiente o la seguridad de los ciudadanos de la UE»[29]. Es claro que los *deep fakes*, siendo representaciones realistas de la realidad, entran fácilmente en la noción de información verificablemente falsa o engañosa. Asimismo, tienen el potencial de causar uno de los perjuicios públicos a los que se refiere el Código de buenas prácticas de la UE: las «amenazas contra los procesos democráticos políticos y de elaboración de políticas».

Uno de los primeros videos que alertaban de esta tecnología, el conocido *deep fake* de Obama creado por Jordan Peele y el medio Buzzfeed, pretendía ilustrar como se podían crear representaciones más o menos convincentes de líderes políticos realizando declaraciones que nunca tuvieron lugar. Un ejemplo más reciente sería el de la guerra de Ucrania, en el que se creó una ultrafalsificación del presidente Zelenski en el que presentaba su rendición frente a las tropas rusas[30]. Por tanto, estas se pueden insertar en el arsenal que diferentes estados utilizan para desestabilizar la opinión pública, si bien su uso se encuentra lejos de ser habitual.

El potencial especialmente dañino que se imputa a las ultrafalsificaciones en este campo parte del argumento del mayor impacto de las representaciones gráficas en los individuos que el texto[31]. Los resultados de algún estudio pionero sugieren que las ultrafalsificaciones sí que pueden aumentar las actitudes negativas contra candidatos que se muestran en esta clase de contenido manipulado[32]. Más aún, su efecto para generar rechazo contra los candidatos o partidos representados en la ultrafalsificación aumenta cuando a través de técnicas de segmentación de audiencias *(microtargeting)* se dirige el contenido falsificado a individuos de

[29] https://digital-strategy.ec.europa.eu/es/library/2018-code-practice-disinformation (consultado 11. 11. 2023).

[30] https://www.elconfidencial.com/tecnologia/novaceno/2022-03-17/hackers-rusos-difunden-un-video-falso-de-zelensky-ordenando-la-rendicion_3393225/ (consultado 11. 11. 2023).

[31] C. Vaccari y A. Chadwick, «Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News», *Social Media + Society*, 6(1). https://doi.org/10.1177/2056305120903408 p. 2.

[32] Véase en este sentido T. Dobber, N. Metoui, D. Trilling, N. Helberger y C. De Vreese, «Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?, *The International Journal of Press/Politics*, 26(1), pp. 69-91. https://doi.org/10.1177/1940161220944364

colectivos muy específicos. Por el contrario, otros estudios sugieren que las ultrasificaciones centradas en políticos necesariamente no son aptas para engañar a los individuos, pero sí para crear una incertidumbre que a la larga puede afectar a la confianza en el ecosistema mediático[33].

Organizaciones como Europol, o fuentes periodísticas reportan también que las ultrafalsificaciones han sido usadas para llevar a cabo varios fraudes. Por ejemplo, el Wall Street Journal reportó que delincuentes utilizaron un programa informático basado en inteligencia artificial, emulando la voz de un ejecutivo para exigir una transferencia fraudulenta[34]. En Turquía, por el contrario, se reporta un caso en el que las ultrafalsificaciones se utilizaron como medio de extorsión[35].

Con todo, el uso más generalizado de las ultra falsificaciones es la creación de pornografía, afectando casi exclusivamente a mujeres que tienen una proyección pública. Y es que, como acertadamente ha señalado Simó Soler «la diversidad que se encuentra en los potenciales usuarios de *deep fakes*, no puede advertirse en quienes las padecen». En este sentido, en un informe del año 2019, la organización Deeptrace extrajo datos de las principales webs de pornografía dedicadas a las ultra falsificaciones, cruzándolos con otras bases de datos de acceso al público, concluyendo que prácticamente el 100% de la pornografía generada por ultrafalsificaciones y otras ediciones más rudimentarias afectaban a mujeres[36].

El Parlamento Europeo considera que las ultrafalsificaciones exacerban las desigualdades de género, convirtiendo a las mujeres en «objetos sin capacidad de defensa»[37], y que suponen una nueva amenaza en este campo. No obstante, lo cierto es que las ultrafalsificaciones pornográficas encajan en un fenómeno más amplio como es la ciber violencia contra las mujeres[38]. Si bien las ultrafalsificaciones más comunes son las que se identifican con pornografía no consentida de celebridades, estas falsificaciones también pueden ser utilizadas para cometer una serie de actos de control sobre la víctima, favorecidas por el entorno tecnológico[39].

En la línea de otras tecnologías de la información mucho más extendidas, las ultrafalsificaciones pueden ser un medio que facilite la

---

[33] C. Vaccari y A. Chadwick, op.cit, p. 9.

[34] https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 (consultado 11.11.2023).

[35] F. Miró Llinares, «Penal law and criminalization in the face of the challenges of AI. General report», *RIDP*, Vol, 1, 2024 (en prensa).

[36] H. Ajder, G. Patrin, F. Cavalli y L. Cullen. «The State of Deepfakes: Landscape, Threats, and Impact», 2019 Disponible en: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf, p. 5.

[37] Parlamento Europeo, *op,cit*, pág. 24.

[38] E. Simó Soler, *op, cit*, p. 498 y ss.

[39] Véase en este sentido P. Lloria García, «La violencia sobre la mujer en el siglo xxi: sistemas de protección e influencia de las tecnologías de la información y la comunicación en su diseño», *La Ley Penal*, Nº 138, Mayo-Junio, 2019, págs. 1-21.

comisión del ilícito. La razón es que el victimario ya no requiere de la captación de la víctima en una situación íntima para llevar a cabo sus propósitos. Previamente, el atentado contra la intimidad debía realizarse bien obteniendo el material sin el consentimiento de la víctima, o con el mismo, pero sin el consentimiento para su divulgación. Ahora, de elegirse este especial medio comisivo, el autor no requiere de ningún material de corte íntimo de la víctima, y los obstáculos que pueda encontrar son el acceso a la tecnología pertinente (el software y el hardware requeridos para realizar la ultra falsificación) y las imágenes de la víctima, que normalmente podrá recopilar de sus redes sociales[40].

No obstante, y sin negar el evidente componente de género que parece indisoluble a las ultrafalsifaciones sexuales, lo cierto es que aún no se reportan casos en los que estas hayan sido utilizadas en esquemas de sextorsión[41], o manifestaciones de la ciber violencia de género[42]. El único supuesto es el mencionado caso de Almendralejo, que no puede calificarse como un caso de violencia de género. Esto seguramente se deba a las dificultades de hacerse con esta tecnología. Como apunta el Research Lab de Europol «dado que los deepfakes se basan en tecnologías avanzadas de IA y aprendizaje automático, se requiere un alto nivel de experiencia para desarrollar la tecnología. En consecuencia, no hay tantos actores de amenazas con el conjunto de habilidades para desarrollarlos por su cuenta». A ello hay que sumar que se necesitan procesadores de alta gama para poder entrenar los modelos en rostros específicos si se quieren conseguir resultados óptimos[43]. Más aún, un alto volumen de imágenes es requerido para entrenar al sistema en una concreta suplantación, si bien algunas aplicaciones parecen estar reduciendo esta necesidad[44].

Eso no significa que la creación de las ultrafalsificaciones no pueda contratarse. Existen mercados en los que se solicitan ultraflsificaciones sexuales , llegando a pagarse sumas de dinero considerables[45]. Hablamos

[40] S., Maddocks, 'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes, *Porn Studies*, 2020, DOI: 10.1080/23268743.2020.1757499 p. 3.

[41] Sobre el término y la casuística jurisprudencial en la materia véase V., Magro Servet, «Consideraciones sobre la delincuencia sexual on line y las víctimas sextorsionadas», *Diario La Ley*, Núm. 9917, Sección Doctrina, 21 de Septiembre de 2021.

[42] R., Rodríguez Fernández, y M. J., Garrido Antón., «Violencia de género a través de internet (ciberviolencia): análisis psicológico-jurídico», *La Ley Penal*, Núm. 154, Sección Estudios, Enero-Febrero 2022.

[43] A., Gómez-de-Ágreda, C., Feijóo, y I., Salazar-García «Una nueva taxonomía del uso de la imagen en la conformación interesada del relato digital. Deep fakes e inteligencia artificial», *Profesional de la informació*n, Vol. 30, 2021, pág. 13.

[44] En este sentido V., Ciancaglini, C., Gibson, C., y D., Sancho, «Malicious Uses and Abuses of Artificial Intelligence», Trend Micro Research, 2020, Disponible en: https://www.europol.europa.eu/cms/sites/default/files/documents/malicious_uses_and_abuses_of_artificial_intelligence_europol.pdf (consultado 11. 11. 2023).

[45] Europol Innovation LAB, *op.cit*, pág.13.

de webs donde los propios usuarios pueden subir un alto número de fotografías para generar el contenido, u otras donde creadores anuncian sus servicios[46]. También existen webs que operan de enlace, perfectamente accesibles desde el buscador de Google donde se aceptan peticiones de videos pornográficos de celebridades y también alojan videos pornográficos creados a través de ultrafalsificaciones. Estas se nutren de los ingresos de la publicidad que se coloca en la página, así como de suscripciones de pago y la promoción de otras aplicaciones. Pese a que sus términos y condiciones prohíben la solicitud de imágenes que no sean de celebridades, repasando muy brevemente los comentarios de estos foros nos damos cuenta de que sirven de punto de encuentro para usuarios y creadores que respectivamente solicitan y proveen videos pornográficos de mujeres a la carta[47].

Asimismo, si el proceso de generación de ultrafalsificaciones más complejas, como un video, encuentra más barreras técnicas, no se dan las mismas condiciones en el caso de imágenes. Este parece ser el supuesto del caso de Almendralejo, en el que los menores utilizaron una aplicación de pago para crear imágenes. Según fuentes periodísticas el modelo utilizado por dicha compañía es una variante de una IA generadora de imágenes, cuyo código fue manipulado para realizar estas ultrafalsificaciones[48].

El previsible aumento de este contenido, de la mano de la «democratización» del software de edición de imágenes y video, ha empujado a adoptar medidas técnicas que pretenden limitar sus efectos perversos. Por un lado, se propone proteger las imágenes que nutren a la IA, por ejemplo, añadiendo patrones de ruido no identificables por el ojo humano que lastran la detección de las caras[49]. También se están haciendo esfuerzos considerables en la detección de ultrafalsificaciones, generalmente a través de sistemas automatizados que tratan de identificar ciertos patrones, con el hándicap que, como hemos apuntado previamente, la lógica de la red generativa adversativa es aprender a burlar esta clase de detectores[50]. Como es lógico, las respuestas legislativas también se están barajando, tanto aquellas que proponen la criminalización de ciertos usos dañinos como las que fijan obligaciones a ciertos actores involucrados en su difusión.

---

[46] H., Ajder, G., Patrin, F., Cavalli, y L., CULLEN, *op.cit*, pág. 5.

[47] Sobre esto con más detalle Kikerpill, K., «Choose your stars and studs: the rise of deepfake designer porn», *Porn Studies*, 2020.

[48] https://www.eldiario.es/tecnologia/negocio-lista-espera-app-usada-desnudar-menores-badajoz-cobra-9-euros-25-fotos_1_10522989.html (consultado 11. 11. 2023).

[49] Parlamento Europeo, *op.cit*, pág. 21.

[50] Europol Innovation LAB, *op.cit*, pág. 13.

3. Propuestas de criminalización de las ultrafalsificaciones

Como se deriva del apartado anterior, las ultrafalsificaciones admiten distintos usos que pueden ser de relevancia para el Derecho penal. En este punto nos referiremos a las dos proposiciones normativas que tienen especial interés para el proceso de criminalización de las ultrafalsificaciones en nuestro ordenamiento jurídico: la propuesta de Directiva sobre la lucha contra la violencia contra las mujeres y la violencia doméstica y la Proposición de Ley Orgánica de regulación de las simulaciones de imágenes y voces de personas generadas por medio de la inteligencia artificial. El análisis, especialmente en lo que concierne al segundo texto elegido[51], se centrará en aquellas disposiciones que sancionan ultrafalsificaciones de carácter sexual, descartándose otros usos que requerirían del análisis de cuestiones previas especialmente conflictivas[52].

Como adelantábamos, la primera de las normativas escogidas es la Propuesta de Directiva del Parlamento Europeo y del Consejo sobre la lucha contra la violencia contra las mujeres y la violencia doméstica. El texto comunitario tiene por objeto prevenir y combatir la violencia contra las mujeres y la violencia doméstica, con el fin de garantizar un alto nivel de seguridad y el pleno disfrute de los derechos fundamentales dentro de la Unión, incluidos el derecho a la igualdad de trato y la no discriminación entre mujeres y hombres. Parte de diferentes instrumentos de *soft law*[53], y de la necesidad de paliar los efectos de la fragmentación legisla-

[51] Hay que tener en cuenta que la Proposición de Ley Orgánica de regulación de las simulaciones de imágenes y voces de personas generadas por medio de la inteligencia artificial pretende introducir un nuevo art. 144 bis en Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General.: «Serán castigados con la pena de tres meses a un año o la de multa de seis a veinticuatro meses quienes desde la convocatoria del proceso electoral y hasta finalizada la jornada de votación difundiesen de forma maliciosa o sin autorización de las personas candidatos afectadas imágenes o audios de estas últimas que estuviesen alterados o recreados mediante sistemas automatizados, software, algoritmos o mecanismos de inteligencia artificial

[52] Me refiero al debate sobre la criminalización de la desinformación, Véase al respecto León Alapont, J., «El Derecho penal ante las fake news y la desinformación: una vuelta de tuerca», *Revista General de Derecho Penal*, 39, 2023, págs 1-59; Devis Matamoros, A., «Criminalización de las fake news en redes sociales: ¿necesidad de intervención o derecho penal simbólico?», *Revista General de Derecho Penal*, Vol. 37, 2022, págs. 1-31.

[53] En este sentido hay que destacar la Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Una Unión de la igualdad: Estrategia para la Igualdad de Género 2020-2025. Pues bien, en ella ya se señalaba que «la inteligencia artificial (IA) se ha convertido en un ámbito de importancia estratégica y un motor clave del progreso económico, por lo que las mujeres deben formar parte de su desarrollo en calidad de investigadoras, programadoras y usuarias. Aunque puede aportar soluciones a muchos retos sociales, se corre el riesgo de que la IA intensifique las desigualdades de género. Si los algoritmos y otros sistemas de aprendizaje automático no son suficientemente transparentes y robustos, existe el riesgo de que se reproduzcan, amplifiquen o alimenten sesgos de género de los que los

tiva en la materia, particularmente antes de la ratificación del Convenio de Estambul por la Unión, que ha experimentado grandes dificultades[54].

La propuesta se fundamenta en el art. 83.1 del Tratado de Funcionamiento de la Unión Europea, que permite al Parlamento Europeo y el Consejo establecer normas mínimas sobre la definición de infracciones penales y sanciones en ámbitos delictivos que sean de especial gravedad y tengan una dimensión transfronteriza, entre ellas los delitos informáticos y la explotación sexual[55]. Pretende abordar fenómenos como la ciberviolencia contra las mujeres, no abordados específicamente en instrumentos supranacionales como el Convenio de Estambul, detectándose «importantes lagunas jurídicas tanto a nivel de la UE como de los Estados miembros»[56]. Así, ante el rápido ritmo que ha cobrado la transformación digital en la actualidad y el aumento de la ciberviolencia la propuesta de Directiva establece unas normas mínimas para determinados delitos informáticos: difusión no consentida de material íntimo o manipulado (artículo 7), delitos relacionados con el ciberacecho (artículo 8), delitos relacionados con el ciberacoso (artículo 9) e incitación al odio o a la violencia por medios cibernéticos (artículo 10).

De esta forma, el art. 7 de la norma empuja a criminalizar en los Estados miembros la difusión no consentida de material íntimo o manipulado[57]. La primera conducta comprendida en el precepto abarca «divulgar

---

programadores no sean conscientes o que son el resultado de una selección de datos específica». No obstante, una cronología de las diferentes iniciativas en la materia, no solo de la UE, puede encontrarse en E., Marí Farinos, «La lucha contra la violencia de género en el derecho comparado, con especial referencia a Europa», *Diario La Ley*, Nº 9128, Sección Tribuna, 29 de enero de 2018, págs. 1-17.

[54] En este sentido, G., kübek «Facing and embracing the consequences of mixity: Opinion 1/19, Istanbul Convention», *Common Market Law Review*, 59(5), 2022, págs. 1465-1500.

[55] La Comisión hace una interpretación amplia del término «explotación sexual» al que se refiere el art. 83.1 del TFUE, pudiendo existir problemas de legitimación normativa en base al Derecho de la Unión. Sobre esto, en detalle, E., Bergamini, «Combating violence against women and domestic violence: From the Istanbul convention to the EU framework: the proposal for an EU directive», *Freedom, Security & Justice: European Legal Studies*, n. 2, 2023, págs. 21-48.

[56] Tampoco la Directiva 2013/40/UE de 12 de agosto de 2013 relativa a los ataques contra los sistemas de información previó las conductas descritas en la medida en que no abordó los llamados cibercrímenes sociales.

[57] La propuesta de Directiva se acerca al modelo de criminalización de las ultrafalsificaciones que se está siguiendo en Reino Unido, de la mano de la Online Safety Bill. Según la Law Commission of England and Wales el ordenamiento jurídico de Reino Unido limitaba los motivos en los que la publicación de estos contenidos era reprimida, particularmente cuando la intención del autor era causar estrés a la víctima (distress), y no abarcando el contenido manipulado (deep fakes). Siguiendo las recomendaciones de la Comisión, el Gobierno de Reino Unido introdujo una enmienda en el procedimiento legislativo de la citada Online Safety Bill, que a su vez enmienda la Sexual Offences Act de 2003. La nueva sección de la citada norma (66 B) castiga la difusión o la amenaza de difundir una fotografía o un video de carácter íntimo. La norma abarca el castigo de las

a una multitud de usuarios finales, mediante el uso de las tecnologías de la información y de las comunicaciones, imágenes, vídeos u otros materiales íntimos que representen actividades sexuales de otra persona sin el consentimiento de esta» (art 7 a)). Nótese que en nuestro ordenamiento esta conducta podría ser castigada a través del delito del art. 197.3 del CP en relación con el art. 197.1, siempre que la información íntima se hubiera obtenido como consecuencia de un acceso no consentido[58]. Por el contrario, cuando no se acredite un apoderamiento no consentido del material íntimo, pero sí una divulgación no consentida, la conducta podrá reconducirse al art 197.7 del CP (STS 693/2023, de 27 de septiembre).

Por el contrario, la letra b) del mismo precepto empuja a los Estados miembros a criminalizar una conducta consistente en «producir o manipular y, posteriormente, divulgar a una multitud de usuarios finales, mediante el uso de las tecnologías de la información y de las comunicaciones, imágenes, vídeos u otros materiales, haciendo que parezca que otra persona está realizando actividades sexuales, sin el consentimiento de esta». Como aclaran los considerandos, esto abarca la fabricación de «ultrafalsificaciones (*deep fakes*), en las que el material se parezca sensiblemente a una persona, a objetos, lugares u otras entidades o acontecimientos existentes, representando actividades sexuales de otra persona, y pueda dar a otros la impresión falsa de que es auténtico o veraz» (considerando 18). Al contrario que la letra a) del artículo 7, esta conducta difícilmente puede encajar en el mencionado art 197.7 del CP, pues el material íntimo no ha sido obtenido de forma consentida de la víctima, sino que ha sido generado a partir de datos que el autor posee y puede haber obtenido de fuentes muy diversas (por ejemplo, de las redes sociales).

La acción típica consiste en «producir o manipular y, posteriormente, divulgar a una multitud de usuarios finales» el contenido. El tipo exige la divulgación del material simulado. Ahora bien, el apartado c) del mismo art. 7 sanciona la amenaza de la difusión del contenido simulado, y «con el fin de coaccionar a otra persona para que realice o acceda a que se realice determinado acto o se abstenga de realizarlo» (art. 7. C), por lo que existe margen para reprimir conductas previas a la difusión. También, respecto a esta última, hay que tener en cuenta que el Committee on Civil Liberties, Justice and Home Affairs y el Committee on Women's Rights and Gender Equality, encargados de crear el borrador de Resolución del Parlamento Europeo que enmendará el texto, han propuesto ampliar el

ultrafalsificaciones a través de una definición amplia del material íntimo, que abarca el material manipulado por sistemas informáticos.

[58] Efectivamente, antes de la difusión del material íntimo debe de haberse realizado algunas alguna de las conductas «invasoras de la intimidad» a las que se refiere el precepto, y en cuya ausencia no es posible apreciar la tipicidad. C., Tomás-Valiente Lanuza, «Delitos contra la intimidad y redes sociales (en especial, en la jurisprudencia más reciente)», en M., Cancio Meliá (ed), *Libro Homenaje al Profesor Dr. Agustín Jorge Barreiro*, UÁM, Madrid, 2019, Vol. 2. pág. 1277.

concepto de difusión, eliminando el requisito de que el contenido manipulado se distribuya a una «multitud» de usuarios finales.

Por otro lado, es necesario que la acción típica se realice mediante el uso de las tecnologías de la información y de las comunicaciones. Estas son definidas por la propuesta de Directiva como todos los recursos y herramientas tecnológicas utilizadas para almacenar, crear, compartir o intercambiar digitalmente información, incluidos los teléfonos inteligentes, los ordenadores, las redes sociales y otras aplicaciones y servicios de los medios de comunicación (art. 4 e)). El papel de estas tecnologías será relevante tanto en la creación del material sintético como en su posterior divulgación. Nótese que la Comisión ha decidido utilizar un concepto amplio del medio comisivo, distinto al de IA, seguramente para evitar posibles divergencias con la regulación aún en tramitación[59]. Esto no será problemático en la medida en que los artefactos tecnológicos abarcados por el término de IA, como decimos, aún en construcción, puedan ser abarcados por categorías existentes, como la que aquí se discute, o la de sistema informático previsto en el Convenio de Budapest[60].

Respecto al contenido de la ultrafalsificación, esta puede consistir en «imágenes, vídeos u otros materiales»[61], siempre que representen a una persona realizando actividades sexuales. La Directiva no define qué debe de entenderse por la realización de actividades sexuales. En la línea con las definiciones aparejadas a los delitos de pornografía infantil, puede defenderse que esto supone la simulación de una persona participando en una conducta sexualmente explícita, en la que también cabe la representación lasciva de los genitales de una persona[62]. Por el contrario, el hecho

---

[59] De hecho, el concepto de ultrafalsificación en este último texto legal es mucho más restringido como veremos *infra*, pues se supedita que la aplicación que genera los contenidos sea considerado IA.

[60] Muy ligado a lo anterior, el borrador de la resolución de la AIDP , Sección II del XXI Congreso en materia de IA y Sistema de Justicia Penal señala: «Some of the criminal laws which punish the production, sale, procurement for use, import, distribution or otherwise making available of devices designed or adapted primarily for the purpose of committing offenses enacted in accordance with articles 2 through 5 of the Budapest Convention can already punish the creation, development and sale of AI systems designed or adapted for those criminal purposes. Thus, as long as AI is considered a device, including a computer program, in accordance with Art. 6 of the Budapest Convention, the introduction of new offences that anticipate the criminal response is not necessary in this field».

[61] El borrador de la resolución del Parlamento Europeo que enmendará el texto, por el contrario, elimina las referencias a las «imágenes, vídeos u otros materiales», sustituyéndolo por el concepto material íntimo. Después, este se define como imágenes, fotografías y grabaciones de vídeo de carácter privado o personal y de naturaleza sexual o que contengan desnudos.

[62] Así, el Informe explicativo sobre el Convenio sobre la Ciberdelincuencia aclara que «La expresión «comportamiento sexualmente explícito» abarca por lo menos las siguientes alternativas, tanto en forma real como simulada: a) las relaciones sexuales, ya sea en forma genital-genital, oral-genital, analgenital u oral-anal, entre menores, o entre un adulto y un menor, del mismo sexo o del sexo opuesto; b) la bestialidad; c) la masturbación; d)

de que la Directiva se refiera expresamente a una actividad sexual hace que estemos ante un concepto más restringido que el tipo de materiales a los que se refiere el art. 197.7 del CP, entre los que sin duda se incluyen las imágenes de contenido sexual pero no exclusivamente (STS 70/2020, de 24 de febrero).

Otro elemento clave es la ausencia del consentimiento de la víctima en la difusión del material. El consentimiento solo es abordado por la propuesta de Directiva al hilo del delito de violación (art. 5). Esta señala que «los Estados miembros se asegurarán de que se entienda por acto no consentido todo acto ejecutado sin el consentimiento voluntario de la mujer o en el que la mujer no pueda formar libremente su voluntad debido a su estado físico o mental, por ejemplo, un estado de inconsciencia, intoxicación, sueño, enfermedad, lesiones corporales o discapacidad, explotando así su incapacidad para formar libremente su voluntad»[63]. Se han escrito ríos de tinta sobre consentimiento en materia de delitos sexuales, en particular tras la LO 10/2022, y no procede aquí una exposición detallada de la problemática generado a su alrededor[64]. No obstante, lo cierto es que el consentimiento (en realidad, «autorización») no ha sido un elemento especialmente conflictivo en el marco del delito del art. 197.7 del CP[65], que contempla un comportamiento de difusión muy similar al que aquí se discute.

---

los abusos sádicos o masoquistas en un contexto sexual, o e) la exhibición lasciva de los genitales o la zona púbica de un menor».

[63] Por su lado, los considerandos señalan que «La falta de consentimiento debe ser un elemento esencial y constitutivo de la definición de la violación, ya que a menudo su perpetración no conlleva violencia física ni uso de la fuerza. El consentimiento inicial debe poder retirarse en cualquier momento durante el acto, de acuerdo con la autonomía sexual de la víctima, y no debe implicar automáticamente el consentimiento para actos futuros».

[64] Por ejemplo, ambos en J. R., Agustina (coord.), *Comentarios a la ley del solo sí es sí. Luces y sombras ante la reforma de los delitos sexuales introducida en la LO 10/2022, de 6 de septiembre*, Atelier, Barcelona, 2023, J. P., Lascuraín Sánchez, «Los nuevos delitos sexuales: indiferenciación y consentimiento», pp. 51-62, t R., Ragués y Vallés, «El grado de afectación al consentimiento de la víctima en los delitos sexuales: una revisión crítica de la Ley Orgánica 10/22», pp. 95-105.

[65] Como es conocido, el debate ha versado sobre otras cuestiones, entre ellas «la oportunidad» de la reforma en atención del «deber de sigilo» que impone a raíz de una expectativa de intimidad que muchos entienden que ha sido previamente desvirtuada por la propia víctima, o el origen de los materiales que en un primer estadio se obtienen de forma consentida (si deben haber sido obtenidos por el victimario o por el contrario es admisible que sea la víctima quién se los facilita). En cambio, como decimos, la falta de autorización posterior en la divulgación del material íntimo no ha sido un elemento cuya apreciación haya sido conflictiva. Véase en este sentido F., Morales Prats, «Delitos contra la intimidad, el derecho a la propia imagen y la inviolabilidad del domicilio», en G., Quintero Olivares (dir.), *Comentarios a la parte especial del Derecho Penal*, Aranzadi, Navarra, 2016, págs. 463 y ss; C., Tomás-Valiente Lanuza, «Delitos contra la intimidad y redes sociales (en especial, en la jurisprudencia más reciente)», en M., Cancio Meliá (ed), *Libro Homenaje al Profesor Dr. Agustín Jorge Barreiro*, *Vol. 2*, UÁM, Madrid, 2019, págs. 1275-1288;

Por otro lado, la Directiva obliga a castigar a los cómplices o inductores de este delito, así como la tentativa del mismo (art. 11). Como es habitual, se impone a los Estados miembros que los comportamientos identificados «se castiguen con sanciones penales efectivas, proporcionadas y disuasorias». En el caso de las conductas previstas en el citado art. 7, que se castiguen con una pena de prisión de una duración máxima de al menos un año (art. 12. 6). Asimismo, se prevén una serie de circunstancias agravantes aplicables también a la difusión no consentida de material íntimo o manipulado, entre ellas, que el delito se haya cometido abusando de una posición reconocida de confianza, autoridad o influencia, o que el delito se haya cometido contra un menor (art. 13).

El segundo texto normativo que va a ser objeto de análisis es la Proposición de Ley Orgánica de regulación de las simulaciones de imágenes y voces de personas generadas por medio de la inteligencia artificial de 13 de octubre de 2023, presentado por el Grupo Parlamentario SUMAR (que, en lo que sigue, abreviaremos como PLO). La PLO pretende modificar la Ley Orgánica 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen, modificando el listado de intromisiones ilegítimas para dar cabida a ciertos usos de las ultrafalsificaciones[66]. También modifica la Ley 13/2022, de 7 de julio, General de Comunicación Audiovisual, introduciendo infracciones que siguen de cerca a la regulación de las ultrafalsificaciones que realiza la propuesta de Reglamento de IA[67].

En lo que a nosotros nos interesa a efectos del análisis, la PLO modifica la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal, in-

[66] Según la propuesta, tendrá la consideración de intromisión ilegítima «La difusión y utilización de imágenes y vídeos de personas o audios de voz generados a través de sistemas automatizados, software, algoritmos o mecanismos de inteligencia artificial sin la previa autorización o consentimiento expreso de la persona o personas afectadas, excepto que incluyan de forma clara y sobresaliente una advertencia de su condición de imagen o audio de voz generado artificialmente por inteligencia artificial. La advertencia deberá figurar sobreimpresa y legible en la propia imagen. Para el caso de los audios de voz deberá realizarse una advertencia audible antes y después de su difusión».

[67] Se incluyen como infracciones muy graves «La difusión de imágenes o vídeos generados a través de sistemas automatizados, software, algoritmos o mecanismos de inteligencia artificial sin la previa autorización o consentimiento expreso de la persona o personas objeto, salvo que incluyan de forma clara y sobresaliente una advertencia de su condición de imagen generada artificialmente por inteligencia artificial. La advertencia deberá figurar sobreimpresa y claramente legible en la imagen, recayendo en caso contrario la responsabilidad sobre la persona que accione dicha generación» y «La difusión de audios generados a través de sistemas automatizados, software, algoritmos o mecanismos de inteligencia artificial que imitan la voz de la persona o personas objeto sin su previa autorización o consentimiento expreso salvo que incluyan de forma clara y perfectamente audible una advertencia de su condición de sonido generado artificialmente por inteligencia artificial. La advertencia deberá realizarse inmediatamente antes y después de haber reproducido la voz generada por estos sistemas, recayendo en caso contrario la responsabilidad sobre la persona que accione dicha generación».

troduciendo un nuevo art. 208 bis. Este último dispone 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». Se acompaña a este nuevo precepto una reforma del art. 211 que estipula que «salvo previa autorización expresa de la persona o personas afectadas, las simulaciones de imágenes, vídeos o audios de voz de estas generados a través de sistemas automatizados, software, algoritmos o mecanismos de inteligencia artificial que fueran difundidos a través de redes sociales serán consideradas como injurias hechas con publicidad».

El primer aspecto relevante a resaltar es que el precepto se introduce en el Título XI del Libro II, relativo a los delitos contra el honor, más concretamente en el Capítulo II dedicado a las injurias. Esto es coherente desde el momento en que la PLO pretende «concretar las conductas injuriosas ya recogidas en el Código Penal» para «dar entrada a las acciones injuriosas en las que para su comisión se utilicen simulaciones de imágenes o voces de personas generadas por inteligencia artificial». El Código Penal define la injuria como la acción o expresión que lesiona la dignidad de otra persona, menoscabando su fama o atentando contra su propia estimación (art. 208). El TS ha señalado que el tipo se encuentra compuesto por un elemento objetivo y otro subjetivo. El elemento objetivo se refiere a que las expresiones o acciones deben constituir un atentado al honor u honorabilidad que sea considerado grave[68]. La valoración de la gravedad deberá realizarse según el Código «en atención a su naturaleza, efectos y circunstancias», existiendo, en palabras del TS, «un relativismo y una enorme circunstancialidad que caracteriza esta infracción» a la hora de valorar y concretar en cada caso lo que socialmente se considera o no grave en este ámbito delictivo (STS 258/2020, de 28 de mayo)[69]. Por otro lado, el elemento subjetivo abarca el «propósito de causar dolor moral con expresiones denigratorias o hirientes para el honor o reputación del sujeto pasivo», que se ha llegado identificar con el tradicional *animus injuriandi*, cuya vigencia como elemento del tipo no es pacífica[70].

[68] Efectivamente, el CP señala que «solamente serán constitutivas de delito las injurias que, por su naturaleza, efectos y circunstancias, sean tenidas en el concepto público por graves, sin perjuicio de lo dispuesto en el apartado 4 del artículo 173». Con la reforma operada por la Ley Orgánica 1/2015, de 30 de marzo, por la que se modifica la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal, se despenalizaron las injurias leves.

[69] En este sentido el TS ha señalado que «a la hora de definir los límites de la tipicidad del delito castigado en el art. 208 del CP, una misma expresión pueda interpretarse, en un determinado contexto, como una interjección coloquial situada extramuros del derecho penal y esa misma palabra, ya en otro entorno, pueda ser valorada como el afilado instrumento para laminar la honorabilidad de un tercero» (STS 669/2022, de 30 de junio).

[70] Por, ejemplo la STS 669/2022, de 20 de junio señala que «En relación con el tipo subjetivo del delito de injurias -»animus iniurandi»- esta Sala ha evolucionado, aunque

Asimismo, y con carácter previo, es necesario delimitar si las expresiones controvertidas se encuentran amparadas por las libertades de expresión e información, de tal forma que la intervención penal no cercene tales libertades (STC 115/2004, de 12 de julio). De esta forma, la propuesta identifica el bien jurídico comprometido con el honor, «concepto jurídico normativo cuya precisión depende de las normas, valores e ideas sociales vigentes en cada momento» pero que protege atentados en la reputación personal, «entendida como la apreciación que los demás puedan tener de una persona, independientemente de sus deseos e impidiendo la difusión de expresiones o mensajes insultantes, insidias infamantes o vejaciones que provoquen objetivamente el descrédito de aquella» (STC 14/2003, de 28 de enero; STS 18/2014, 17 de enero)[71].

El contenido simulado debe ser recreado «mediante sistemas automatizados, software, algoritmos o inteligencia artificial». Pudiera parecer, que al contrario que la Propuesta de Directiva sobre la lucha contra la violencia contra las mujeres y la violencia doméstica, el tipo es más restrictivo en lo que atañe a los medios para generar los contenidos,

---

no siempre con la uniformidad que hubiera sido deseable, en la exigencia de ese elemento subjetivo del injusto que históricamente operaba como una suerte de dolo especial o reduplicado. Y es que, frente al previgente art. 453 del CP que definía el delito de injurias como «toda expresión proferida o acción ejecutada en deshonra, descrédito o menosprecio de otra persona», el actual art. 208 no enfatiza ya el elemento tendencial que se advierte en la primera de las definiciones. El precepto ahora vigente admite, sin dificultad alguna, que la injuria pueda ser cometida con dolo directo o con dolo eventual». En detalle sobre esta cuestión J.L., Fuentes Ossorio, «Elementos subjetivos en los delitos contra el honor», *Estudios Penales y Criminológicos*, vol. XXIX, 2009, pp. 271-310.

[71] Señala Muñoz Conde que el concepto de honor es determinado por un elemento objetivo, la fama o reputación social, y uno subjetivo, la propia estimación, y que «ambos son tenidos en cuenta en el art. 208 como determinantes del concepto de injuria que se da en dicho precepto». F. Muñoz Conde, *Derecho Penal. Parte Especial*, Tirant lo Blanch, Valencia, 2023, p. 330. Por su lado, Sánchez Ostiz entiende el honor como un bien jurídico funcional, «al servicio de la protección de otros nucleares o intrínsecos». A su vez, matiza que «el honor se encuentra conformado por la percepción de la víctima respecto a lo que su entorno entiende de ella; en pocas palabras, la percepción por la víctima de la valoración que de ella tiene el entorno. Por tanto, se configura gracias a la confluencia a) de la percepción social del entorno, y b) la percepción de a) por la víctima» lo que le lleva a señalar que «la injuria requiere cierto volumen de difusión en el contexto o entorno para poder apreciar la tipicidad (en este sentido entiendo que el tipo básico sería deseable exigir cierta entidad por la publicidad de las afirmaciones, de manera que no sea este la modalidad agravada)». P., Sánchez Ostiz, «Los delitos contra el honor: una propuesta de sistematización», en D. M., Santana Vega (dir.), Una *perspectiva global del Derecho Penal, Libro homenaje al profesor Dr. Joan J. Queralt Jiménez,* Atelier, España, 2021 pp. 805-816. Miró Llinares por su parte apuesta, de forma complementaria a la cuestión del bien jurídico, por importar el principio de ofensa de la tradición del *common law*, para deslindar aquellas lesiones del honor que no deben ser reprimidas por el Derecho Penal y sí otras ramas del ordenamiento como el Derecho Civil, así como reservar a las conductas dañinas (*harm*) la pena de prisión, excluyéndose en el caso de las ofensas. F. Miró Llinares, «Injuriar es ofender: apuntes sobre la criminalización de los delitos contra el honor desde el enfoque teórico del daño/ofensa», en M., Cancio Meliá (ed), *Libro homenaje al profesor Dr. Agustín Jorge Barreiro*, AUM, Madrid, 2019, págs. 1161-1174.

especialmente al citarse la IA. Por el contrario, al referirse también a conceptos genéricos como software o algoritmo nos encontramos en el mismo supuesto que el tipo previsto en la norma europea, esto es, ante una indefinición de los medios tecnológicos utilizados para recrear la imagen o la voz.

También comparte, con matices, otros elementos. La recreación debe realizarse sin la autorización de la víctima. Asimismo, la recreación de la imagen corporal o la voz de la víctima debe realizarse «para la pública difusión». Si bien la intención de difundir la ultrafalsificación es un aspecto relevante, elemento subjetivo del tipo junto al ánimo de menoscabar el honor, no se requiere que la difusión tenga lugar. En otras palabras, la recreación no autorizada de la imagen o de la voz es suficiente para la consumación. La difusión, cuando el contenido se propague a través de algunos de los medios previstos (la imprenta, la radiodifusión o por cualquier otro medio de eficacia semejante), determinará la aplicación de la modalidad agravada de injurias a la que se refiere el art. 211 del CP.

La diferencia más relevante entre ambas propuestas de reforma es que la PLO es mucho más amplia en lo que se refiere a las características del contenido generado. Recordemos que en la Directiva el objeto de la ultrafalsificación es una representación de un sujeto realizando actividades sexuales, mientras que la PLO es indiferente ante el carácter de la representación, siempre que esta sea apta para lesionar el honor de la víctima. Por tanto, si bien lógicamente una ultrafalsificación que represente a una persona realizando un acto sexual sería reprimida por el tipo propuesto, también serían subsumibles en el tipo otros supuestos. Pienso en la representación de la voz de una persona que simule unas declaraciones que nunca tuvieron lugar, punible cuando se aprecie la lesión al honor, operación en la que habrá que tener especialmente en cuenta el juego que pueden tener otras libertades constitucionales[72].

A mi juicio, esto supone una inconsistencia con la propia exposición de motivos de la norma. Haciéndose eco de muchos de los trabajos ya señalados en el apartado segundo de este mismo trabajo, se señala que las ultrafalsificaciones «tienen un significativo sesgo de género, ya que el 90 por cierto suplantan la identidad de mujeres, lo que, además de la vulneración de derechos que supone en sí misma, aumenta el riesgo de incidencia de los casos de acoso y otras formas de violencia sobre la mu-

[72] La propia exposición de motivos admite que «El uso de la inteligencia artificial está amparado bajo la libertad de expresión recogida en el artículo 20.1 de la Constitución española, cuando reconoce el derecho a la producción y la creación artística, así como el de «expresar y difundir libremente los pensamientos, ideas y opiniones mediante (...) cualquier otro medio de reproducción»". De hecho, la propuesta de Reglamento de IA, cuando se refiere a ciertos deberes de los operadores de ultrafalsificaciones, prevé excepciones para aquellos contenidos que se compartan en el ejercicio de derecho a la libertad de expresión y el derecho a la libertad de las artes y de las ciencias (art. 52.3).

jer» y que «este tipo de actividades debe ser entendido como una nueva y peligrosa forma de violencia sexual». En este sentido, es criticable que este sesgo de género, o manifestaciones de las ultrafalsificaciones especialmente vejatorias como las que implican una recreación de la mujer realizando actos sexuales, no hayan sido objeto de tratamiento específico. La respuesta penal prevista es idéntica en todos los supuestos, y no atiende adecuadamente a la gravedad de la recreación cuando esta tiene por objeto un aspecto tan íntimo como es el propio cuerpo, manipulado de forma grotesca para satisfacer los deseos sexuales de otras personas o el interés económico del creador de la ultrafalsificación.

Muy ligado a lo anterior, el tipo contemplado en la propuesta de Ley Orgánica realiza una aclaración bienvenida pero prescindible[73]. El tipo de injurias no exige un medio determinado para su comisión. El propio enunciado del tipo se refiere a una dualidad entre acciones y expresiones, lo que indica que las injurias «pueden realizarse valiéndonos de un lenguaje diverso al hablado, a través, por ejemplo, de caricaturas, alegorías o emblemas»[74]. Lo relevante es que «se infrinja un deber de comportarse aceptado por la comunidad y que ello se considere objetivamente como injurioso»[75]. El concreto medio comisivo para realizar dicho atentado no es ningún impedimento para afirmar la tipicidad.

Y es que también es pacífico que las injurias pueden difuminarse por distintos medios técnicos, lo cual es respaldado por el propio Código cuando se refiere a la imprenta, la radiodifusión o cualquier medio de eficacia semejante a efectos de aplicación del art. 211. De hecho, las nuevas tecnologías de la información han servido como canal para difundir contenidos injuriosos, de tal forma que la afectación al honor puede ser incluso mayor[76]. Igualmente, es ilustrativo que cuando el *sexting* de tercero no se encontraba tipificado en el art. 197.7 del CP, los tribunales

---

[73] Es el caso de una enmienda propuesta al Código penal Italiano. En esta propuesta, los deep fakes se definen como «cualquier imagen y/o vídeo, realizado de cualquier forma, que combine y/o superponga imágenes y/o vídeos de una persona sobre otras imágenes y/o vídeos de otra persona, con el fin de generar un vídeo realista, pero falso». Además, la ley expresa que se entenderá por falsificaciones profundas aquellos contenidos que sean «publicados y difundidos sin el consentimiento y autorización de la persona a la que se refiere el video «falso», vulnerando sus derechos a la intimidad, honor, imagen, decoro, reputación y cualquier otro derecho». La ley castiga a quien, sin el consentimiento de la persona afectada, cree y envíe o entregue, difunda o publique, por cualquier medio, un «deep fake» de los contemplados en el artículo 1, salvo que el hecho constituya un delito más grave. La enmienda objeto de examen puede consultarse aquí en italiano: https://www.senatoragazzi.it/iniziative/disegno-di-legge/103/

[74] A. Pablo Serrano, *Honor, Injurias y Calumnias. Los delitos contra el honor en el Derecho histórico y en el derecho vigente español*, Tirant lo Blanch, Valencia, 2018, pág. 255.

[75] Muñoz Conde, *Derecho Penal. Parte Especial, op.cit*, pág. 332.

[76] En detalle R., Miguel Barrio, «El delito de injurias y las redes sociales. El número de 'followers' y otras variables ambientales como elementos de valoración del daño», *Revista de Internet, Derecho y Política*, 2022, n.º 36, págs. 1-13.

acudían al tipo de injurias para castigar este comportamiento[77]. En este sentido, existen claras similitudes entre ambos comportamientos, pues en ambos media una difusión no consentida de un material que sin duda tiene la capacidad de lesionar el honor de la víctima[78].

En cualquier caso, la PLO, de aprobarse paralelamente con la Directiva, será insuficiente en lo que se refiere al marco penológico. La PLO equipara la pena al delito de injurias, que recordemos son castigadas con de multa de seis a catorce meses, cuando son realizadas con publicidad, y en el tipo básico con la de tres a siete meses (art. 209). Pues bien, la Directiva obliga a que las conductas descritas sean castigadas «con sanciones penales efectivas, proporcionadas y disuasorias», siendo muy discutible que este sea el caso. Más aún, la exclusiva previsión de una pena de multa para las conductas referidas al nuevo art. 208 bis viola la literalidad del art. 12.6 de la propuesta de Directiva que señala que «los Estados miembros se asegurarán de que las infracciones penales a que se refieren los artículos 7 y 9 se castiguen con una pena de prisión de una duración máxima de al menos un año». Consiguientemente, es previsible que sean necesarias reformas adicionales en el caso de que ambas proposiciones legislativas prosperen de forma paralela[79].

4. Obligaciones extrapenales a operadores de ultrafalsificaciones y su relevancia penal

4.1. Planteamiento general

Como ha puesto de manifiesto Miró Llinares, la prevención de ilícitos cometidos a través de la IA puede suponer que la respuesta penal

[77] Véase en este sentido A., Mendo Estrella, «Delitos de descubrimiento y revelación de secretos: acerca de su aplicación al *sexting* entre adultos», *Revista Electrónica de Ciencia Penal y Criminología,* núm. 18-16, 2016, págs. 1-27; E., Núñez Castaño, «La relevancia penal de las nuevas tecnologías y su incidencia en los denominados ciberdelitos: especial referencia a los delitos contra la intimidad», *Revista General de Derecho Penal*, Núm. 37, 2022.

[78] Eso no quita que pueda argumentarse que las ultrafalsificaciones sexuales lesiones otros bienes jurídicos, cuestión aun escasamente desarrollada por la doctrina. A estos efectos véase el análisis preliminar de A. Devis Matamoros, *Algunas claves del castigo penal del deepfake de naturaleza sexua*l, Iberconect, 2023, Disponible en: https://www.ibericonnect.blog/2023/07/algunas-claves-del-castigo-penal-del-deepfake-de-naturaleza-sexual/ (última consulta 20. 1. 2024).

[79] Si esto ocurre no debe descartarse que se reconduzca la represión de las ultrafalsificaciones que consisten en la creación de pornografía no consentida al art. 197, añadiendo un nuevo apartado al precepto, tal como ya ha apuntado algún autor. En realidad, esta opción, no exenta de controversia respecto al problema del bien jurídico, sería más cercana al texto de la Directiva, que regula conjuntamente la difusión de material íntimo de forma no consentida con la difusión de ultrafalsificaciones que simulan a una persona realizando actividades sexuales. Véase E., Simó Soler, *op.cit*.

deba anticiparse a momentos previos a la externalización del daño. Esto se justifica porque en muchos casos es en estos momentos previos, la producción del sistema, su diseño, donde el daño puede prevenirse, debiendo dirigirse el Derecho penal a regular la conducta de aquellos, que, en estas fases, pueden efectivamente reducir el riesgo de lesión de los bienes jurídicos afectados[80].

Hasta el momento las propuestas de criminalización discutidas no han prestado atención a este aspecto en el caso de las ultrafalsificaciones, y parecen estar ideadas para regular la conducta de aquellos que diseminan el contenido. Por el contrario, existen otras normativas administrativas, algunas ya en vigor, que establecen obligaciones a los sujetos involucrados en el ciclo de vida de estos sistemas, y que posibilitan la creación y diseminación de las ultrafalsificaciones.

Como apunta el borrador de la resolución de la Sección II de la AIDP en materia de Inteligencia Artificial y Derecho penal, los debates sobre la transformación del Derecho Penal por el impacto de la IA no pueden pasar por alto los debates relativos a su regulación en otras ramas del ordenamiento jurídico[81]. El estudio de la normativa extrapenal, de los intereses protegidos y de los comportamientos sancionados por aquella, es necesario para adoptar la compleja decisión de a qué instrumento represivo le compete la tutela de los daños creados por la IA[82].

---

[80] F., Miró Llinares, «Penal law and criminalization in the face of the challenges of AI. General report», *op.cit*, pág. 17 y ss. El borrador de resolución en la Sección II del XXI Congreso Internacional de Derecho Penal también incluye una resolución específica en esta línea «Criminal law systems are designed to have a deterrent effect on likely offenders, preventing them from engaging in criminal actions. If the key moment in terms of risk in relation to AI is the moment of the design and implementation, the enactment of offenses that aim to deter conducts at such moments shall be considered. This can be done by anticipating the protection with endangerment offenses, that punish not following certain duties in relation to specific interests worthy of protection. Also, and similar to what is established with the criminal liability of legal persons, specific regulatory obligations related to the design and implementation of AI systems could be established, which infringement may give rise to criminal liability».

[81] «The development of AI may give rise to new interests worthy of protection. Additionally, AI systems can affect the dimension and relevance of interests that are not currently considered worthy of protection by Criminal law. When criminal laws do not provide an adequate response to protect these interests, new criminal offenses shall be enacted that proportionally punish conducts that are harmful to such interests. This shall only be done when there are no alternative means that are less harmful than Criminal law to effectively protect the mentioned interests».

[82] La resolución del XIV Congreso Internacional de Derecho penal, titulada «Los problemas jurídicos y prácticos planteados por la diferencia entre Derecho penal y Derecho penal administrativo (Sección I)» señala «Whether or not certain conduct should properly be punished according to criminal law or to administrative penal law cannot be determined categorically. lt is therefore in most cases for the legislature to decide what conduct is to be sanctioned criminally or by administrative penal law. In making that decision, legislatures should take into consideration several criteria, especially the importance of the social

Más aún, la normativa administrativa puede identificar una serie de comportamientos antijurídicos, cuyas manifestaciones más graves pueden, en su caso, constituir objeto de sanción penal. No se trata de que de la infracción del deber previsto en la norma administrativa se derive automáticamente la responsabilidad penal[83]. Más bien, esta normativa puede ayudar a identificar qué comportamientos van más allá del riesgo permitido[84], algo que puede ser muy relevante en supuestos como en el de los procesos algorítmicos, donde puede haber muchos sujetos implicados[85].

En el tema que nos ocupa, el texto normativo que puede ayudarnos a identificar estos deberes es la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley De Inteligencia Artificial). Como decíamos, esta es una normativa aún en tramitación y que está experimentado cambios sustanciales[86], también en lo relativo a la regulación de las ultrafalsificaciones[87]. El objetivo del Reglamento «es promover en la Unión la adopción de una inteligencia artificial fiable y centrada en el ser humano y garantizar un elevado nivel de protección de la salud, la seguridad, los derechos fundamentales, la democracia y el Estado de Derecho y del medio ambiente frente a los efectos nocivos de los sistemas de inteligencia artificial, apoyando al mismo tiempo la innovación y mejorando el funcionamiento del mercado interno». Hay que tener en cuenta que

---

interest affected by the conduct in question, the gravity of endangerment or harm to that interest, and the kind and degree of fault on the part of the offender».

[83] Sobre esto recientemente Silva Sánchez sobre las conductas neutras de los prestadores de servicios intermediarios. J. M., Silva Sánchez, «¿Hasta qué punto son conductas neutras los servicios de Google, Facebook o Twitter?», *InDret*, 3.2023, págs. 1-3.

[84] Siguiendo de nuevo a Silva Sánchez, la normativa extrapenal puede crear de forma *ex ante* un espacio de libertad jurídica de acción que ampara a ciertos sujetos generadores de peligro, esto es, un haz de comportamientos que reciben el nombre de «riesgo permitido» y que ha sido encajado en el edificio de la Teoría del Delito de diversas formas. J. M., Silva Sánchez, *El riesgo permitido en el Derecho penal económico*, Atelier, Barcelona, 2022, p. 35.

[85] A., Moraiti, «AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation», *RIDP*, 2021, pág. 109 y ss. En un sentido similar, Picotti, entre las distintas soluciones que recoge en su informe para clarificar la cuestión del *responsability gap* por la IA cuando no se aprecia el dolo, y se materializa un daño, una de ellas es que el legislador identifique áreas de riesgo permitido, determinando el uso social adecuado de la IA, L., Picotti, *op.cit*, p. 43.

[86] Advertimos que trabajamos con la última versión enmendada por el Parlamento Europeo, Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html

[87] Tras la aceptación de este trabajo se ha conocido el texto final del Reglamento de Inteligencia Artificial, afectando a algunas de las obligaciones de los operadores de ultrafalsificaciones. Particularmente, la categoría referente a los modelos fundacionales ha sido eliminada, trasladándose parte de su regulación a los modelos de IA de uso general y los modelos de IA de uso general que presentan riesgos sistémicos. Asimismo, se han introducido precisiones muy relevantes relativas a los sujetos a los que están dirigidos ciertas obligaciones de transparencia, paliando ciertas carencias de la regulación.

el Reglamento establece «un marco jurídico uniforme en el desarrollo, la introducción en el mercado, la puesta en servicio y la utilización de la inteligencia artificial». Esto es muy relevante pues garantiza la libre circulación transfronteriza de bienes y servicios basados en la IA, impidiendo «que los Estados miembros impongan restricciones al desarrollo, la comercialización y la utilización de sistemas de inteligencia artificial (sistemas de IA)» a menos que el Reglamento lo autorice expresamente (considerando 1).

En el ciclo de vida de la IA, y no es el caso distinto de las ultrafalsificaciones, intervienen distintos sujetos con diferentes roles y deberes. A efectos de aclarar responsabilidades Valls Prieto propone una división de sujetos responsables dependiendo de la fase en la que actúan y el tipo de control que ejercen sobre el sistema inteligente. Así, distingue entre: desarrolladores y fabricantes del producto que lo diseñan y mandan al mercado; profesionales que utilizan estos sistemas inteligentes para realizar una parte de su trabajo y los usuarios finales[88]. Por el contrario, la propuesta de Reglamento de Inteligencia Artificial establece un esquema de obligaciones administrativas mucho más complejo. Este se dirige a quienes define globalmente como operadores, que abarca categorías muy distintas de sujetos implicados en el ciclo de la IA, como los proveedores, importadores, distribuidores e implementadores (usuarios de IA profesionalizados). Asimismo, el Parlamento Europeo en el proceso legislativo pretende introducir en el texto categorías adicionales como los sistemas de IA basados en licencias libres y de código abierto o los modelos fundacionales. En aras de facilitar la exposición de la regulación extrapenal de las ultrafalsificaciones diferenciaremos, en primer lugar, una serie de obligaciones impuestas a los operadores de la IA, distinguiendo en su seno las diferentes subcategorías. Después haremos una especial mención a ciertas obligaciones específicas que se dirigen a los usuarios de las ultrafalsificaciones (denominados implementadores). Finalmente nos ocuparemos de otros agentes no regulados por el Reglamento, que también están sujetos a obligaciones para mitigar sus efectos adversos en la normativa de servicios digitales.

4.2. Obligaciones de los operadores de ultrafalsificaciones

Como decíamos la propuesta de Reglamento de Inteligencia Artificial (versión enmendada por el Parlamento) utiliza el término operador para aglutinar a una serie de sujetos involucrados en el ciclo de la IA, entre

[88] J., Valls Prieto, J., «Sobre la responsabilidad penal por la utilización de sistemas inteligentes», *Revista Electrónica de Ciencia Penal y Criminología*, 24-27, 2022, págs. 1-35.

ellos los proveedores de los sistemas[89]. Con carácter preliminar hay que señalar que hay ciertos operadores que no se encuentran sujetos al Reglamento: aquellos operadores de modelos IA de Código abierto. Como hemos apuntado previamente, sistemas que han servido para crear ultrafalsificaciones estaban basados en este tipo de modelos, por lo que cabría pensar, que, como primer eslabón en la cadena, estos operadores deberían estar sujetos a algún tipo de obligación para prevenir subsiguientes ilícitos. La versión del Reglamento enmendada por el Parlamento presta especial a estos sistemas de código abierto, señalando que «a fin de impulsar el desarrollo y el despliegue de la inteligencia artificial, especialmente entre las pymes, las empresas emergentes y la investigación académica, pero también entre los particulares» (considerando 12 bis) el Reglamento no se aplicará a los componentes de IA proporcionados en el marco de licencias libres y de código abierto, «salvo en la medida en que sean comercializados o puestos en servicio por un proveedor como parte de un sistema de IA de alto riesgo o de un sistema de IA incluido en el ámbito de aplicación de los títulos II o IV» (artículo Artículo 2. 5 sexies)[90].

Únicamente deberá alentarse a los desarrolladores de componentes de inteligencia artificial libres y de código abierto «a que apliquen prácticas de documentación ampliamente adoptadas, como modelos y tarjetas de datos, como una forma de acelerar el intercambio de información a lo largo de la cadena de valor de la inteligencia artificial, permitiendo la promoción de sistemas de IA fiables en la Unión» (considerando 12 bis). Por el contrario, la fijación de deberes específicos de conducta se ha descartado, justamente en atención a los beneficios que estos desarrollos tecnológicos aportan al campo. Eso induce a pensar que los riesgos que pueden darse dentro de estos estadios previos del desarrollo deben entenderse, al menos de forma general, dentro de un «riesgo permitido» y consiguientemente, pese a que se materialice un daño al que causalmente pueden haber contribuido, estos sujetos no deben ser responsables del mismo.

Despejada esta cuestión inicial, procede identificar a qué obligaciones están sujetos los operadores ultrafalsificaciones sujetos al Reglamento. En sus distintas versiones, el Reglamento clasifica a los sistemas de IA en base a su riesgo, aparejando una serie de obligaciones en función

[89] Hay que tener en cuenta que el proveedor es a efectos del Reglamento «toda persona física o jurídica, autoridad pública, agencia u organismo de otra índole que desarrolle un sistema de IA o para el que se haya desarrollado un sistema de IA con vistas a introducirlo en el mercado o ponerlo en servicio con su propio nombre o marca comercial, ya sea de manera remunerada o gratuita».

[90] En este sentido el Reglamento aclara que «Ni el desarrollo colaborativo de componentes de inteligencia artificial libres y de código abierto ni su puesta a disposición en repositorios abiertos deben constituir actividades de comercialización o puesta en servicio» (considerando 12 ter).

del mismo a los operadores, y prohibiendo ciertas prácticas de IA (art. 5). Si atendemos al caso de las ultrafalsificaciones, estas no se clasifican como prácticas prohibidas[91], ni tampoco como sistemas de alto riesgo, habiendo sido descritas como sistemas de riesgo limitado[92]. Esto supone que operadores e implementadores no deben someterse a los requisitos del Capítulo II y II del Título II (mantenimiento de sistema de gestión de riesgos, obligaciones relativas a la precisión, solidez y ciberseguridad, entre otros). Por el contrario, los operadores solo deberán tener en cuenta los principios generales aplicables a todos los sistemas de IA, introducidos por las enmiendas del Parlamento (art. 4 bis), y los que tengan la consideración de implementadores, como veremos en el apartado siguiente, lo dispuesto en el art. 53.3.

No obstante, los operadores de ultrafalsificaciones pueden encontrarse sujetos a obligaciones adicionales en la medida en el que el sistema de IA entre dentro de una categoría específica: los modelos fundacionales. Esta ha sido introducida por la última modificación del texto por el Parlamento, en aras de regular «avances recientes» en la IA de la mano de estos modelos, «existiendo una incertidumbre significativa sobre el modo en que evolucionarán los modelos fundacionales, tanto en lo que se refiere a la tipología de los modelos como a su autogobernanza» (considerando 60 octies). Los modelos fundacionales son definidos como «un modelo de sistema de IA entrenado con un gran volumen de datos, diseñado para producir información de salida de carácter general y capaz de adaptarse a una amplia variedad de tareas diferentes» (art. 3. 1 quater). En lo que aquí nos interesa, el Reglamento distingue una subcategoría de estos modelos como es la IA generativa, que define indirectamente como «sistemas de IA destinados específicamente a generar, con distintos niveles de autonomía, contenidos como texto, imágenes, audio o vídeo complejos («IA generativa»)». Pues bien, los proveedores de modelos fundacionales y los proveedores que especialicen un modelo fundacional en un sistema de IA generativo deberán cumplir las obligaciones de transparencia establecidas en el artículo 52, apartado 1[93], esto es, garantizarán que los sistemas de IA destinados a interactuar con personas físicas estén diseñados y desarrollados de forma «que el sistema de IA, el propio proveedor o el usuario informen de manera clara, inteligible y

[91] Eso no quita que otra legislación del Derecho de la Unión pueda prohibir ciertas prácticas de IA (art. 5.1 bis versión del Parlamento). Por el contrario, es muy dudoso que los Estados miembros puedan prohibir prácticas de IA, por ejemplo, criminalizando, sin el amparo de una norma comunitaria, contrariando el Reglamento y consiguientemente el Derecho de la Unión.

[92] http://euro.ecom.cmu.edu/program/law/08-732/AI/EU-AIAct.pdf

[93] Nótese que la remisión es únicamente al apartado 1 del artículo 52, y no a su apartado tercero, claramente el Reglamento está pensando en este caso a grandes modelos de lenguaje de texto y no IA generativa de imágenes.

oportuna a dichas personas físicas expuestas a un sistema de IA de que están interactuando con un sistema de IA».

Más aún, los proveedores «formarán y, en su caso, diseñarán y desarrollarán el modelo fundacional de manera que se garanticen salvaguardias adecuadas contra la generación de contenidos que infrinjan el Derecho de la Unión, en consonancia con el estado de la técnica generalmente reconocido y sin perjuicio de los derechos fundamentales, incluida la libertad de expresión». El Reglamento empuja pues a introducir medidas preventivas para prevenir que la IA genere contenidos ilícitos en base a las peticiones de los usuarios. Esto ya puede apreciarse en los estándares de la industria. Modelos de lenguaje como ChatGPT han establecido un sistema de moderación de contenidos para valorar *input* del usuario (el texto que sirve como indicación para el modelo generativo), y así evitar que ofrezca resultados sobre indicaciones clasificadas previamente como discurso del odio, contenido sexual o que puede empujar a la autolesión[94]. Esto también es aplicable a sistemas generativos entrenados para crear imágenes siguiendo las instrucciones de los usuarios, que podrían ser utilizados para crear ultrafalsificaciones[95]. No obstante, y tal como apuntan expertos contactados por Europol, existen vulnerabilidades evidentes del sistema de moderación que empujan a su mejora constante[96].

Es perfectamente posible pues, que normativas administrativas, o incluso la Ley penal, establezca sanciones pertinentes para que los proveedores de IA generativa mantengan sistemas de control de riesgos que minimicen su posible uso con propósitos delictivos. Este modo de proceder no es nada novedoso, y se asemeja a modelos de responsabilidad penal por el producto, solución que por otra parte la doctrina ya ha esbozado para mitigar ciertas lagunas punitivas en el campo de la IA[97]. En cualquier caso, si nos adentramos en un modelo de responsabilidad penal por el producto, en el que la normativa administrativa tiene un importante papel en la determinación de la conducta típica, habrá que clarificar los requisitos mínimos de lesividad[98] para evitar las tensiones

[94] Open AI, «New and improved content moderation tooling», 2023, https://openai.com/blog/new-and-improved-content-moderation-tooling (consulta 11. 11. 2023).

[95] Pensemos en el caso del modelo DALE 2 también suministrado por Open AI, https://help.openai.com/en/articles/6338764-are-there-any-restrictions-to-how-i-can-use-dall-e-2-is-there-a-content-policy (consulta 11. 11. 2023).

[96] Europol, «ChatGPT The impact of Large Language Models on Law Enforcement», 2023, Disponible en: https://www.europol.europa.eu/media-press/newsroom/news/criminal-use-of-chatgpt-cautionary-tale-about-large-language-models (consulta 11. 11. 2023).

[97] S., Aires de Sousa, «Portuguese report on traditional criminal law categories and AI», *RIDP*, Vol. 94 issue 1, 2023, págs. 319-330.

[98] F., Muñoz Conde, «La responsabilidad por el producto en el derecho penal español», *Derecho & Sociedad*, Núm. 49, 2017, págs. 253-279.

con los principios que deben informar al Derecho penal en un Estado Democrático y de Derecho[99].

4.3. Obligaciones dirigidas a los implementadores de sistemas de ultrafalsificaciones: la caótica regulación del deber de información del contenido sintético y su posible relevancia penal

Como decíamos, el Reglamento distingue entre los operadores de sistemas de IA a los implementadores, esto es, «toda persona física o jurídica, autoridad pública, agencia u organismo de otra índole que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional». La normativa establece obligaciones específicas para los implementadores[100] de ultrafalsificaciones, reguladas en el art. 52.3. Ante riesgos «específicos de suplantación o falsificación» que acarrean estos sistemas (considerando 70) el mencionado art. 52.3 establece que los implementadores «harán público de manera adecuada, oportuna, clara y visible que el contenido ha sido generado de forma artificial o manipulado, así como, cuando sea posible, el nombre de la persona física o jurídica que lo generó o manipuló». La norma específica que «por hacer público» se entenderá el etiquetado del contenido de un modo que se informe que no es auténtico, y que resulte claramente visible para su destinatario, «debiendo de tenerse en cuenta el estado de la técnica generalmente reconocido y las normas y especificaciones armonizadas pertinentes a esos efectos». Dicha información se facilitará a más tardar con ocasión de la primera interacción o exposición, debiendo de ser «accesible a las personas vulnerables, como las personas con discapacidad o los niños, y se completará, cuando proceda y sea apropiado, con procedimientos de intervención o de denuncia para la persona física expuesta teniendo en cuenta el estado de la técnica generalmente reconocido, las normas armonizadas pertinentes y las especificaciones comunes» (art. 53. 3 ter). No obstante, la norma establece una excepción al etiquetado para aquellos sistemas que estén «legalmen-

[99] Una discusión muy rica puede encontrarse en C., Grandi, «Positive obligations (*garantestellung*) grounding criminal responsibility for not having avoided an il-legal result connected to the Ai functioning», *RIDP*, Vol. 94 issue 1, 2023, págs. 67-77.

[100] En realidad, el texto del Parlamento mantiene el término usuario en este punto, lo cual no se entiende muy bien puesto que este ha sido sustituido por el de implementador en las definiciones. Hay que tener en cuenta que la definición que hacía la Comisión de «usuario» en su propuesta coincide con la de implementador :»toda persona física o jurídica, autoridad pública, agencia u organismo de otra índole que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional». Así, la enmienda 172 del Parlamento se limita a sustituir un término por otro sin alterar la definición.

te autorizados por la ley» o su uso resulte necesario para el ejercicio del derecho a la libertad de expresión y el derecho a la libertad de las artes y de las ciencias[101].

Como decíamos, dicha obligación está dirigida a los implementadores. Esto puede presentar problemas pues solo abarca a quien hace uso del sistema de ultrafalsificación, esto es, quién en último término obtiene la imagen del sistema para después compartirla con terceros y no a quién pone a disposición al público el sistema generativo (proveedor o distribuidor dependiendo del caso). Debemos recordar que los implementadores sujetos al Reglamento son aquellos que utilizan la IA en el ámbito profesional (art. 3. 4). En otros términos, usuarios que exploten servicios como *Face swap* de Google o *Dall e 2* de Open AI no estarán sujetos a esta obligación.

Por otro lado, esta obligación no incumbe al proveedor, esto es el encargado del desarrollo del sistema. Esta parece una decisión meditada considerando que el apartado 1 del mismo art. 53 sí establece como sujeto destinatario de la obligación al proveedor, mientras que en este caso se habla del usuario (implementador). De esta forma, quien facilite el aplicativo no deberá colocar la etiqueta, siendo sus únicos deberes los relativos a los sistemas fundacionales generativos, en los términos ya descritos[102]. Desde esta perspectiva es muy extraño que la norma empuje a establecer «procedimientos de intervención o de denuncia para la persona física expuesta teniendo en cuenta el estado de la técnica generalmente reconocido» si el obligado es un implementador y no un proveedor[103].

Por tanto, en la medida en la que el Reglamento utiliza un concepto restringido de implementador (aquellos que usan la IA con propósitos profesionales) y se excluye a los proveedores de esta obligación, los sujetos obligados se reducen considerablemente. En medio de esta confusión, y con una precipitación del todo innecesaria a la regulación de IA, la PLO enmienda la Ley 13/2022, de 7 de julio, General de Comunicación Audiovisual, castigando una conducta muy similar a la descrita. Así, se prevé como infracción muy grave «la difusión de imágenes, vídeos o audio generados a través de sistemas automatizados, software, algoritmos o mecanismos de inteligencia artificial sin la previa autorización o consentimiento expreso de la persona o personas objeto, salvo que incluyan

[101] En esta línea «cuando el contenido forme parte de una obra cinematográfica claramente creativa, satírica, artística o ficticia, de imágenes de videojuegos y de obras o formatos análogos, las obligaciones de transparencia establecidas en el apartado 3 se limitarán a revelar la existencia de tales contenidos generados o manipulados de una manera adecuada, clara y visible, que no obstaculice la presentación de la obra, y a revelar los derechos de autor aplicables, cuando proceda» (art. 53. 3 bis).

[102] Recordemos que a este último solo le son aplicables las obligaciones relativas a los sistemas generativos (art. 28, ter).

[103] No es extraño que el texto final se dirija ahora a los proveedores de sistemas de ultrafalsificaciones y haya eliminado la referencia a estos procedimientos.

de forma clara y sobresaliente una advertencia de su condición de imagen generada artificialmente por inteligencia artificial o en el audio». La proposición se presenta como un desarrollo enormemente deficiente del Reglamento de IA, que, al no identificar el operador específico que debe cumplir con tal obligación incluye como infracción administrativa una conducta que no constituye un incumplimiento del Reglamento.

El cumplimiento del deber de etiquetado puede llegar a tener incidencia en la responsabilidad penal, pues, desde el momento en el que el contenido se presenta al público como falso, la lesión al bien jurídico puede entenderse como menos intensa. Esto podría tener alguna virtualidad cuando la ultrafalsificación se utiliza en el contexto de una parodia; pensemos en el uso no consentido de la voz de una persona con relevancia pública para crear un contenido que se haga viral en Internet. No obstante, difícilmente podrá alegarse que el etiquetado convierte en atípica la conducta cuando el contenido simulado es de carácter sexual. La afectación al honor no reside en sembrar una duda razonable sobre si el comportamiento de la víctima tuvo o lugar, sino en la propia utilización de la imagen de la persona para realizar una representación que es de por sí denigrante.

4.4. Obligaciones de los prestadores de servicios intermediarios en el Reglamento de Servicios Digitales

La amenaza que presentan las ultrafalsificaciones no se detiene en la creación del contenido simulado y preocupa especialmente su posterior propagación a través de distintos canales de difusión. Nos referimos evidentemente a plataformas en línea de muy gran tamaño, por ejemplo, redes sociales como Facebook, X o TikTok, pero también a prestadores de alojamiento de datos dedicados a ofrecer contenidos pornográficos. Asimismo, como hemos apuntado en el apartado segundo desde este mismo trabajo, estas webs pueden ser la génesis del contenido ultrafalsificado, en el sentido de que sirven de punto de encuentro para quien dispone de los medios técnicos para generar las imágenes simuladas y quien está dispuesto a sufragar su creación. Desde un punto de vista preventivo, el posicionamiento que adopten estos prestadores de servicios de la sociedad de la información frente a las ultrafalsificaciones será relevante, pues permitirá reducir su propagación, o, incluso, impedir que estas se creen en primer lugar. Lógicamente, implicados en el proceso de difusión de las ultrafalsificaciones, procede preguntarse qué deberes establece la legislación frente a los prestadores de servicios de la sociedad de la información, y cuál es su relación con una eventual responsabilidad penal.

La Directiva de Comercio Electrónica, traspuesta en nuestro ordenamiento a través de Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico, guió el debate sobre la res-

ponsabilidad penal de los intermediarios, estableciendo amplias exenciones de responsabilidad que limitaban el campo de aplicación de la ley penal a los intermediarios. La Directiva ha sido recientemente reemplazada por el Reglamento 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022, relativo a un mercado único de servicios digitales (en lo que sigue DSA por sus siglas en inglés), que mantiene en líneas generales el sistema de la Directiva con varias novedades muy relevantes que afectan a las ultrafalsificaciones.

Respecto a la responsabilidad de los intermediarios de alojamiento de datos, el tipo de intermediario que nos interesa especialmente a efectos del estudio, la DSA sigue estableciendo el requisito del conocimiento ilícito del prestador como pilar para apreciar la responsabilidad del intermediario (art. 6 DSA). La novedad reside en que la DSA establece procedimientos de notificación y acción, que permiten a usuarios denunciar la presencia de contenidos ilícitos alojados en el servicio del prestador (art. 16). De esta forma, la presencia de interfaces para canalizar denuncias es obligatoria para todos los prestadores dedicados a alojar datos de terceros, y con independencia de su tamaño. Las notificaciones dirigidas al prestador a través de estos canales, siempre que cumplan con unas formalidades específicas (art. 16.2), son un instrumento apto para atribuir al prestador el conocimiento efectivo del ilícito a efectos del art. 6. La DSA precisa que esto ocurrirá cuando el prestador pueda determinar «sin un examen jurídico detallado», que la información o la actividad pertinentes son ilícitas (art. 16.3)[104]. Todo ello permite a los usuarios denunciar la presencia de ultrafalsificaciones en estos servicios, enervándose la exención de responsabilidad si el prestador no retira los contenidos ilícitos cuando sea consciente de su presencia.

Asimismo, dicho conocimiento es un elemento clave para afirmar la responsabilidad del prestador en base a los tipos penales previstos en el Derecho nacional. Así, por ejemplo, Gómez Tomillo defendía respecto a la Ley 34/2002, de 11 de julio, argumento extrapolable a la DSA, que, fijando determinados deberes en la remoción de contenidos, la normativa extrapenal constituye la base de la posición de garantía de los prestadores de servicios a efectos de declarar su responsabilidad por comisión por omisión[105]. De forma análoga, Galán Muñoz entiende que estas exenciones crean un ámbito de riesgo permitido cuando los prestadores de servicios se comportan de la forma fijada en las normas de comercio elec-

[104] Este es un requisito muy similar al fijado en la jurisprudencia en el ámbito civil en la interpretación de la Ley 34/2002, de 11 de julio. En este sentido véase A., Soler Presas, ¿»Am I in Facebook? Sobre la responsabilidad civil de las redes sociales on-line por la lesión de los derechos de la personalidad, en particular por usos no consentidos de la imagen de un sujeto», *InDret*, 3/2011, págs. 1-44.

[105] Señala que en estos supuestos estaríamos ante M. Gómez Tomillo, *Responsabilidad penal y civil por Delitos Cometidos a través de Internet*, Aranzadi, Navarra, 2006, p. 127 y ss.

trónico, integrándose su apreciación en el juicio de tipicidad[106]. Aunque no puede identificarse conocimiento efectivo del ilícito con el dolo, sin duda este último puede ser un indicio claro de que concurre el elemento subjetivo del tipo, algo que los nuevos procedimientos de notificación y acción facilitarán[107].

Ahora bien, el aporte más relevante en materia de regulación de las ultrafalsificaciones que ofrece la DSA es su explícita inclusión en el sistema de gestión de riesgos que crea la norma en su Capítulo III, Sección V. De forma muy resumida, la DSA establece unas obligaciones, denominadas de diligencia debida, cuyo objetivo es garantizar un entorno en línea transparente y seguro. Empuja a los grandes prestadores de plataformas en línea y motores de búsqueda a desplegar medidas para gestionar riesgos sistémicos que se derivan del funcionamiento de sus servicios[108]. Entre estos riesgos se encuentra la diseminación de contenidos ilícitos, entre los que pueden incardinarse contenidos sexuales ilícitos como las ultrafalsificaciones (considerando 80)[109]. Compete a plataformas y motores de muy gran tamaño evaluar y gestionar estos riesgos, fiscalizándose estas actividades por diferentes medios fijados en el Capítulo IV de la norma.

En lo que se refiere a las medidas de mitigación, el Reglamento introduce ciertas de carácter facultativo, entre las que se menciona expresamente a las ultrafalsificaciones. Los prestadores de muy gran tamaño deberán de «garantizar que un elemento de información, ya se trate de imagen, audio o vídeo generado o manipulado que se asemeja notablemente a personas, objetos, lugares u otras entidades o sucesos existentes y que puede inducir erróneamente a una persona a pensar que son auténticos o verídicos, se distinga mediante indicaciones destacadas cuando se presente en sus interfaces en línea y, además, proporcionar una funcionalidad fácil de utilizar que permita a los destinatarios del servicio

---

[106] En detalle sobre la espinosa relación entre las exenciones de responsabilidad y la Teoría General del Delito A. Galán Muñoz, *Libertad de expresión y responsabilidad penal por contenidos ajenos en internet, Un estudio sobre la incidencia penal de la ley 34/2002 de Servicios de la Sociedad de la Información y el Comercio Electrónico*, Tirant monografías, Valencia, 2010, pág. 68 y ss.

[107] Miró Llinares, F., «Cibercriminalidad y responsabilidad de los prestadores de servicios a la luz de la normativa europea y de su interpretación por los Tribunales españoles», en Gómez Martín, V, (coord.), *Garantías constitucionales y Derecho penal europeo*, Marcial Pons, España, págs. 561-584.

[108] Sobre el nebuloso concepto de riesgo sistémico en la DSA S., Broughton Micova, S y A., Calef, «Elements for effective systemic risk assessment under the DSA», CERRE, 2023, Disponible en: https://cerre.eu/publications/elements-for-effective-systemic-risk-assessment-under-the-dsa/ (última consulta 22.1.2024)

[109] En realidad, el Reglamento, como norma horizontal y dirigida a establecer normas generales que afectan a la responsabilidad y deberes de las plataformas utiliza un concepto de contenido ilícito amplio, que es una remisión a la normativa especial de la Unión, y a la de los Estados miembros.

señalar dicha información» (art. 35. 1. k)). Consiguientemente, la DSA empuja a introducir sistemas de etiquetado de las ultrafalsificaciones, de forma análoga a la regulación en materia de IA. Todo ello empuja a entender que la difusión de las ultrafalsificaciones puede ser un riesgo sistémico en el marco del Reglamento y el etiquetado de las mismas en aras de determinar su carácter sintético una medida de reducción de riesgos.

El cumplimiento de estas obligaciones de diligencia debida se asegura por un complejo sistema de ejecución de corte administrativo en el que se establecen amplias facultades de ejecución a los coordinadores de servicios digitales y a la Comisión Europea, entre ellas, la imposición de multas en función del volumen de negocio. La relación entre estas normas y la responsabilidad de los intermediarios es oscura, ya que ambos aspectos son disociados expresamente por el Reglamento[110]. Esto no debe entenderse problemático pues estas nuevas obligaciones, más que un medio para afirmar la responsabilidad del prestador, constituyen una alternativa preventiva a la responsabilidad para llevar a cabo la labor de tutela de los bienes jurídicos afectados[111].

En esta línea, ya es habitual que los grandes intermediarios proscriban o limiten en sus condiciones la diseminación de ultrafalsificaciones[112]. Esto afecta a prestadores como Facebook o Instagram, pero también a grandes servicios de difusión de pornografía, entre ellos Pornohub, XVideos o Stripchat, que de hecho recientemente han sido declarados por la Comisión como prestadores de plataforma en línea de muy gran tamaño y obligados a gestionar los mencionados riesgos sistémicos de la DSA. Queda por ver si esta novedosa labor de mitigación de riesgos es ejecutada de forma efectiva.

5. Breves conclusiones

La criminalización de determinados usos de las ultrafalsificaciones es una operación compleja que requiere tener en cuenta tanto factores

[110] Dispone la DSA en su considerando 40 que «Las obligaciones de diligencia debida son independientes de la cuestión de la responsabilidad de los prestadores de servicios intermediarios y, por tanto, deben apreciarse por separado».

[111] «Las obligaciones de diligencia debida son independientes de la cuestión de la responsabilidad de los prestadores de servicios intermediarios y, por tanto, deben apreciarse por separado» (considerando 41). Véase en este sentido Miguel Asensio, P. M., «Obligaciones de diligencia y responsabilidad de los intermediarios: el Reglamento (UE) de Servicios Digitales», *La Ley Unión Europea*, Núm. 109, 2022, págs. 1-48; Husovec, M., «Rising above liability: the Digital Services Act as a blueprint for the second generation of global internet rules», *Berkeley Technology Law Journal*, Vol. 38., 2023, págs. 101-137.

[112] META, *Enforcing Against Manipulated Media*, 6 de enero de 2020, Disponible en: https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/ (última consulta 20 de enero de 2024).

normativos como criminológicos. Por lo que se refiere a estos últimos, parece claro que las ultrafalsificaciones pueden utilizarse para afectar a bienes jurídicos de máxima relevancia, y que el sesgo de género de las ultrafalsificaciones es patente. No obstante, aún se reportan pocos supuestos en los que esta tecnología se ha utilizado con propósitos delictivos, lo cual posiblemente se deba a la complejidad del entrenamiento de estos sistemas, que dificulta su acceso al público general. Esto evidencia que no es necesaria una acción urgente del legislador, siendo aconsejable que la criminalización de las ultrafalsificaciones se plantee cuando las distintas iniciativas a nivel de la Unión hayan prosperado.

Y es que, desde un punto vista normativo la coexistencia de diferentes instrumentos normativos, tramitada de forma paralela, puede dar lugar a disfunciones que empañan la de por sí compleja decisión de qué comportamientos criminalizar. Más aún, esta respuesta normativa se tramita a la vez que una regulación de corte administrativo que está imponiendo deberes de conducta de cara a mitigar de ciertos riesgos aparejados a las ultrafalsificaciones. Estos pueden ser muy relevantes para identificar áreas de riesgo permitido que no deberían interesar al Derecho penal (el caso de los operadores de sistemas de IA de Código abierto, excluidos en el Reglamento de IA). También, estas normativas, y el régimen sancionador que las acompaña, pueden ser un instrumento suficiente para tutelar los bienes jurídicos comprometidos, sin que sea necesaria la intervención del Derecho penal (pienso en el sistema de gestión de riesgos que impone el Reglamento de Servicios Digitales). Todo ello empuja a la cautela en un momento en que los procesos de criminalización planteados en el ámbito nacional no pueden ser ajenos a la regulación comunitaria, muy relevante más allá de propuestas de criminalización como la mencionada propuesta de Directiva sobre la lucha contra la violencia contra las mujeres y la violencia doméstica

6. Bibliografía

Abadía Selman, A., «La protección penal de los trabajadores frente a la inteligencia artificial en el ámbito del delito de discriminación laboral», *Revista General de Derecho Penal*, 39, 2023, págs. 1-61.

Aires De Sousa, S., «Portuguese report on traditional criminal law categories and AI», *RIDP*, Vol. 94 issue 1, 2023, págs. 319-330.

Ajder, H., Patrin, G., Cavalli, F. y Cullen, L., «The State of Deepfakes: Landscape, Threats, and Impact», 2019 Disponible en: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf.

Bergamini, E., «Combating violence against women and domestic violence: From the Istanbul convention to the EU framework: the proposal for an EU directive», *Freedom, Security & Justice: European Legal Studies*, n. 2, 2023, págs. 21-48.

Broughton Micova, S. y Calef, A., «Elements for effective systemic risk assessment under the DSA», CERRE, 2023, Disponible en: https://cerre.eu/publications/elements-for-effective-systemic-risk-assessment-under-the-dsa/.

Ciancaglini, V., Gibson, C. y Sancho, D., «Malicious Uses and Abuses of Artificial Intelligence», Trend Micro Research, 2020, Disponible en: https://www.europol.europa.eu/cms/sites/default/files/documents/malicious_uses_and_abuses_of_artificial_intelligence_europol.pdf.

Citron, D. y Chesney, R., «Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security», *California Law Review*, 197, 2019, págs.1769-1820.

Comisión Europea, «Libro blanco sobre la inteligencia artificial. Un enfoque europeo orientado a la excelencia y la confianza», 2020, Disponible en: https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065&from=ES.

De La Mata Barranco, N. J., «Reflexiones sobre el bien jurídico a proteger en el delito de acceso informático ilícito (art. 197 bis cp)», *Cuadernos de política criminal*, Núm. 118, 2016, págs. 43-86.

Devis Matamoros, A., «Criminalización de las fake news en redes sociales: ¿necesidad de intervención o derecho penal simbólico?», *Revista General de Derecho Penal*, Vol. 37, 2022, págs. 1-31.

— *Algunas claves del castigo penal del deepfake de naturaleza sexual*, Iberconect, 2023, Disponible en: https://www.ibericonnect.blog/2023/07/algunas-claves-del-castigo-penal-del-deepfake-de-naturaleza-sexual/.

Dobber, T., Metoui, N., Trilling, D., Helberger, N. y De Vreese, C., «Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?, *The International Journal of Press/Politics*, 26(1), págs. 69-91. https://doi.org/10.1177/1940161220944364.

Europol Innovation Lab, «Facing reality? Law enforcement and the challenge of deepfakes», 2022, disponible en: https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes.

Europol, «ChatGPT The impact of Large Language Models on Law Enforcement», 2023, Disponible en: https://www.europol.europa.eu/media-press/newsroom/news/criminal-use-of-chatgpt-cautionary-tale-about-large-language-models.

Galán Muñoz, A., «La internacionalización de la represión y la persecución de la criminalidad informática: un nuevo campo de batalla en la eterna guerra entre prevención y garantías penales», *Revista Penal*, Núm. 24, 2009, págs. 90-107.

— *Libertad de expresión y responsabilidad penal por contenidos ajenos en internet. Un estudio sobre la incidencia penal de la ley 34/2002 de Servi-*

cios de la Sociedad de la Información y el Comercio Electrónico, Tirant monografías, Valencia, 2010.

Gómez Tomillo, M., *Responsabilidad penal y civil por Delitos Cometidos a través de Internet*, Aranzadi, Navarra, 2006.

Grandi, C., «Positive obligations (*garantestellung*) grounding criminal responsibility for not having avoided an il-legal result connected to the Ai functioning», *RIDP*, Vol. 94 issue 1, 2023, págs. 67-77.

Hacker, P., Engel, A. y Mauer, M., «Regulating ChatGPT and other Large Generative AI Models», en ACM Conference on Fairness, Accountability, and Transparency (FAccT '23), June 12–15, 2023, págs. 1123-1223.

High Level Expert Group On Fake News And Online Disinformation, «A multi-dimensional approach to disinformation», 2018, Disponible en: https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation.

Husovec, M., «Rising above liability: the Digital Services Act as a blueprint for the second generation of global internet rules», *Berkeley Technology Law Journal*, Vol. 38. 2023, págs. 101-137.

Ice, J., «Defamatory Political Deepfakes and the First Amendment», *Case W. Rsrv. L. Rev*, Vol. 417, 2019, págs. 417-455.

Juanatey Dorado, C. «Protección penal de la intimidad frente a la utilización ilícita de medios digitales. Un análisis de la reciente doctrina jurisprudencial», *Revista General de Derecho Penal*, 39, 2023, págs. 1-45.

Kübek, G., «Facing and embracing the consequences of mixity: Opinion 1/19, Istanbul Convention», *Common Market Law Review*, 59(5), 2022, págs. 1465-1500.

León Alapont, J., «El Derecho penal ante las fake news y la desinformación: una vuelta de tuerca», *Revista General de Derecho Penal*, 39, 2023, págs. 1-59.

Lloria García, P., «Delitos y redes sociales: los nuevos atentados a la intimidad, el honor y la integridad moral». Especial referencia al «sexting», *La Ley Penal*, Nº 105, Sección Estudios, Noviembre-Diciembre, 2013, págs. 1-10.

— «La violencia sobre la mujer en el siglo xxi: sistemas de protección e influencia de las tecnologías de la información y la comunicación en su diseño», *La Ley Penal*, nº 138, mayo-junio, 2019, págs. 1-21.

— «La difusión de imágenes íntimas sin consentimiento (A propósito de la Sentencia 70/2020 del Tribunal Supremo de 24 de febrero de 2020)», *LA LEY privacidad* , Nº 4, Sección El foro de la privacidad, Segundo trimestre de 2020, págs. 1-7.

Maddocks, S., «A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political' deep fakes», *Porn Studies*, 7:4, 2020, págs. 415-423.

Marí Farinos, E., «La lucha contra la violencia de género en el derecho comparado, con especial referencia a Europa», *Diario La Ley*, Nº 9128, Sección Tribuna, 29 de enero de 2018, págs. 1-17.

Mendo Estrella, A., «Delitos de descubrimiento y revelación de secretos: acerca de su aplicación al sexting entre adultos», *Revista Electrónica de Ciencia Penal y Criminología*, núm. 18-16, 2016, págs. 1-27.

Miguel Asensio, P. M., «Obligaciones de diligencia y responsabilidad de los intermediarios: el Reglamento (UE) de Servicios Digitales», *La Ley Unión Europea*, Núm. 109, 2022, págs. 1-48.

Miguel Barrio, R., «El delito de injurias y las redes sociales. El número de 'followers' y otras variables ambientales como elementos de valoración del daño», *Revista de Internet, Derecho y Política*, N.º 36. 2022, págs. 1-13.

Miró Llinares, F., «Penal law and criminalization in the face of the challenges of AI. General report», RIDP, Vol, 2, 2024 (en prensa).

— «Cibercriminalidad y responsabilidad de los prestadores de servicios a la luz de la normativa europea y de su interpretación por los Tribunales españoles», en Gómez Martín, V, (coord.), *Garantías constitucionales y Derecho penal europeo*, Marcial Pons, España, 2012, págs. 561-584.

— «Injuriar es ofender: apuntes sobre la criminalización de los delitos contra el honor desde el enfoque teórico del daño/ofensa», en M., Cancio Meliá (ed), *Libro homenaje al profesor Dr. Agustín Jorge Barreiro*, UAM, Madrid, 2019, págs. 1161-1174.

— «Inteligencia artificial, delito y control penal: nuevas reflexiones y algunas predicciones sobre su impacto en el derecho y la justicia penal», *El Cronista del Estado Social y Democrático de Derecho*, Vol. 100, 2022, págs. 174-183.

Moraiti, A., «AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation», *RIDP*, 2021, págs. 109-122.

Muñoz Conde, F., «La responsabilidad por el producto en el derecho penal español», *Derecho & Sociedad*, Núm. 49, 2017, págs. 253-279.

— *Derecho Penal. Parte Especial*, Tirant lo Blanch, Valencia, 2023.

Núñez Castaño, E., «La relevancia penal de las nuevas tecnologías y su incidencia en los denominados ciberdelitos: especial referencia a los delitos contra la intimidad», *Revista General de Derecho Penal*, Núm. 37, 2023, págs. 1-45.

Pablo Serrano, A., *Honor, Injurias y Calumnias. Los delitos contra el honor en el Derecho histórico y en el derecho vigente español*, Tirant lo Blanch, Valencia, 2018.

Palma Herrera, J. M., «Inteligencia artificial y ciencias penales. Aproximación a las bases de una compleja relación», en Galán Muñoz, A., y Mendoza Calderón, S., (coord.), *Derecho penal y política criminal en tiempos convulsos*, Tirant lo Blanch, Valencia, 2021, págs. 37-70.

Parlamento Europeo, «*Tackling deepfakes in European policy*», 2021, Disponible en: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2021)690039.

Picotti, L., «Traditional Criminal law categories and AI: crisis or palingenesis? General report», *RIDP*, Vol. 94 issue 1, 2023, págs. 11-53.

Rodríguez Fernández, R. y Garrido Antón, M. J., «Violencia de género a través de internet (ciberviolencia): análisis psicológico-jurídico», *La Ley Penal*, Núm. 154, Sección Estudios, Enero-Febrero, 2022, págs, 1-11.

Sánchez-Ostiz, P., *A vueltas con la Parte Especial (Estudios de Derecho Penal)*, Atelier, Barcelona, 2020.

Silva Sánchez, J. M., «¿Hasta qué punto son conductas neutras los servicios de Google, Facebook o Twitter?», *InDret*, 3.2023, págs. 1-3.

— *El riesgo permitido en el Derecho penal económico*, Atelier, Barcelona, 2022.

Simó Soler, E., «Retos jurídicos derivados de la Inteligencia Artificial Generativa Deepfakes y violencia contra las mujeres como supuesto de hecho», *InDret*, 2.2023, págs. 1-23.

Soler Presas, A., ¿»Am I in Facebook? Sobre la responsabilidad civil de las redes sociales on-line por la lesión de los derechos de la personalidad, en particular por usos no consentidos de la imagen de un sujeto», *InDret*, 3/2011, págs. 1-44.

UK Centre For Data Ethics And Innovation, «Snapshot Paper - Deepfakesand Audiovisual Disinformation», 2020, Disponible en: https://www.gov.uk/government/publications/cdei-publishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai/snapshot-paper-deepfakes-and-audiovisual-disinformation.

Vaccari, C. y Chadwick, A., «Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News», *Social Media + Society*, 6(1). 2020, págs. 1-23.

Valls Prieto, J., «Sobre la responsabilidad penal por la utilización de sistemas inteligentes», *Revista Electrónica de Ciencia Penal y Criminología*, 24-27, 2022, págs. 1-35.

Van Der Sloot, B. y Wagensveld, Y., «Deepfakes: regulatory challenges for the synthetic society», *Computer Law & Security Review*, Volume 46, 2022, págs. 1-15.

Zittrain, J., «The Generative Internet», *Harvard Law Review*, Vol. 119, 2006, págs. 1975-2023.