InDret
2.2023

Elisa Simó Soler
Universitat de València
Universidad de Buenos Aires

Retos jurídicos derivados de la Inteligencia Artificial Generativa

Deepfakes y violencia contra las mujeres como supuesto de hecho

Sumario

El desarrollo de nuevas aplicaciones de Inteligencia Artificial requiere de un estudio sobre su impacto social. En concreto, invita a realizar análisis prospectivos de las consecuencias jurídico-políticas que tendría el empleo de herramientas de generación automática de contenido audiovisual hiperrealista personalizado. Profundizar en el binomio deepfakes-violencia contra las mujeres conlleva la identificación de tres categorías problemáticas de empleo de contenido audiovisual fraudulento en contra de las mujeres a través de: a) técnicas de desprestigio, b) suplantación de identidad y c) simulación de situaciones constitutivas de delito. Frente a los potenciales usos perjudiciales, se articula un sistema de triple respuesta que interpela al ámbito jurídico, político y tecnológico desde la multidisciplinariedad y la interdependencia para formular propuestas de solución desde la vigencia de los derechos humanos de las mujeres y los valores democráticos.

Abstract

The development of new Artificial Intelligence applications requires a study of their social impact. In particular, it invites prospective analyses of the legal-political consequences of using tools to automatically generate personalized hyper-realistic audiovisual content. The deepfakes-violence against women binomial involves the identification of three problematic categories of use of fraudulent audiovisual content against women through a) smear techniques, b) impersonation and c) simulation of situations constituting a crime. In the face of the potentially harmful purposes, a triple response system is articulated that challenges the legal, political, and technological spheres from a multidisciplinary and interdependent perspective to formulate proposals for solutions based on the validity of women's human rights and democratic values.

Title: *Legal challenges arising from Generative Artificial Intelligence: Deepfakes and violence against women as a factual scenario.*

Palabras clave: Inteligencia Artificial Generativa, Deepfakes, Violencia contra las mujeres, Suplantación de identidad, Engaño, Posverdad.
Keywords**:** *Generative Adversarial Networks, Deepfakes, Violence Against Women, Identity Theft, Deception, Post-truth.*

DOI: 10.31009/InDret.2023.i2.11

InDret

2.2023

Recepción
09/01/2023
-
Aceptación
31/03/2023
-

Índice

-

-

Este trabajo se publica con una licencia Creative Commons Reconocimiento-No Comercial 4.0 Internacional

1. Introducción: nuevos desarrollos de Inteligencia Artificial y de violencia contra las mujeres*

Es ya una obviedad, y podría rozar la reiteración absurda, hacer alusión a la incorporación en la vida diaria de las personas de sistemas de Inteligencia Artificial (IA). No obstante, pese a la naturalización (casi inconsciente) de muchos de los dispositivos de uso rutinario, los continuos avances en este campo de la ciencia de la computación abren nuevos interrogantes que interpelan directamente a las ciencias jurídicas. Cada uno de los progresos debe ser evaluado a la luz de los principios y derechos presentes en los ordenamientos jurídicos y se deben cuestionar las garantías y riesgos que el empleo de IA conlleva en cada supuesto concreto[1].

Los ritmos acelerados de la evolución tecnológica, en compás asincrónico con el desarrollo normativo, exigen al Derecho la agudeza de plantear análisis prospectivos que permitan anticipar los vacíos legales o articular las herramientas jurídicas que eviten prácticas que colisionen con los derechos fundamentales, al tiempo que doten de cobertura legal a aquellas que refuerzan un estándar proteccionista de esos mismos derechos[2]. La relación tecnología-derecho puede leerse en términos de prestación de asistencia mutua con el objetivo de suplir las respectivas carencias. En palabras de FRANGANILLO, «la misma tecnología capaz de resolver viejos problemas puede traer problemas nuevos si no se utiliza de forma adecuada»[3]. La capacidad resolutiva que se le otorga a la IA puede reportarse del Derecho que tendrá que ser capaz de ofrecer respuestas legales a las nuevas incógnitas que los modelos de IA puedan generar. Esto es, la IA asiste para solucionar problemas pasados y el Derecho responde para evitar conflictos futuros.

1.1. ¿Inteligencia Artificial al servicio de las *deepfakes*?

La manipulación de fotografías, textos, audios o vídeos con múltiples finalidades no es un fenómeno nuevo, pero sí lo es, en cambio, el proceso de realización. Lo que anteriormente precisaba de un ejercicio manual con una importante inversión de tiempo, ahora se genera a través de lo que se conoce como redes generativas adversariales (*generative adversarial networks* o GAN por sus siglas en inglés). Se trata de una innovación del aprendizaje automático que permite la creación de contenidos audiovisuales hiperrealistas y personalizados de forma instantánea.

Los sistemas GAN se componen de dos modelos (generativo y discriminativo) que son funcionalmente adversarios y que GOODFELLOW et al. describen con los roles de falsificador y policía. El modelo generativo se correspondería con un equipo de falsificadores que intentan

* Elisa Simó Soler (elisa.simo@uv.es). Este trabajo se ha realizado en el marco de las ayudas Margarita Salas para la formación de jóvenes doctores del programa de recualificación del sistema universitario español, financiado por el Ministerio de Universidades del Gobierno de España y la Unión Europea (Next Generation EU). Asimismo, la publicación forma parte de la investigación iniciada en el seno del Proyecto de Generación del Conocimiento «Claves para una justicia digital y algorítmica con perspectiva de género» (PID2021-123170OB-I00) del que soy integrante como investigadora posdoctoral.

[1] SURDEN, «Artificial intelligence and law: An overview», *Georgia State University Law Review*, núm. 35, 2019, pp. 1335-1337.

[2] BARONA VILAR, «Inteligencia Artificial o la algoritmización de la vida y de la justicia: ¿solución o problema?», *Rev. Boliv. de Derecho*, núm. 28, 2019, p. 45 y NIEVA FENOLL, Jordi, *Inteligencia artificial y proceso judicial*, Marcial Pons, Madrid, 2018.

[3] FRANGANILLO, «Contenido generado por inteligencia artificial: oportunidades y amenazas», *Anuario ThinkEPI* 16, 2022, p. 2.

producir dinero falso, mientras que el modelo discriminativo se personificaría en los cuerpos policiales. La competencia entre ambos, que constituye la fase de entrenamiento, conduce a una falsificación cada vez más realista hasta que, finalmente, la policía no es capaz de distinguir entre moneda real y falsa[4].

Aplicado al ámbito de la imagen, el resultado que ofrecen las GAN es la producción de fotografías o vídeos falsos de alta fidelidad. Esta técnica de síntesis de imágenes basada en IA recibe el nombre de *deepfake* (traducido al castellano como *ultrafalso*), término compuesto por la palabra *deep* (proveniente de Deep Learning o aprendizaje profundo) y *fake* (falso en inglés) y prestado del nombre de un usuario de *Reddit* que en 2017 publicó vídeos pornográficos haciendo uso de esta tecnología con actrices de Hollywood como Taylor Swift o Scarlett Johansson[5].

Encontrándose en una fase experimental, las posibilidades de creación de contenido parecen infinitivas: síntesis de imágenes de caras nuevas (no preexistentes), suplantaciones de identidad (intercambiando la cara de una persona por otra), manipulación de atributos (edición de cara o retoques: color de piel, pelo, ojos, edad, género...), cambios de expresión (recreación facial), sincronización del movimiento de los labios con un discurso, reproducción de movimientos o adición de filtros en tiempo real en videoconferencias[6]. Así ha sido posible devolver a la vida a personajes como Salvador Dalí en el museo de San Petersburgo en Estados Unidos generando una experiencia inmersiva e interactiva con el artista y su obra[7] o recrear a Lola Flores en el anuncio «Con mucho acento» de Cruzcampo y a Luis Aragonés para publicitar la Liga finalizando el spot con la frase confirmatoria de la longevidad digital que permiten estas nuevas técnicas de IA «cuando haces algo grande, vives para siempre».

Como se puede intuir, los usos y funcionalidades que se pueden derivar de las *deepfakes* son muy diversas y, en muchos de los supuestos, positivas. Cabe su incorporación en el ámbito de la medicina, la educación, la industria, el arte, la publicidad, el entretenimiento o el marketing[8]. Sin embargo, los avances podrían instrumentalizarse con fines que atenten directamente contra

---

[4] GOODFELLOW ET AL., «Generative adversarial networks», *Communications of the ACM*, vol. 63, núm. 11, 2020, p. 141 y KORSHUNOV, Pavel y SÉBASTIEN, Marcel, «Deepfakes: a new threat to face recognition? Assessment and detection», 2018, p. 2 arXiv preprint arXiv:1812.08685

[5] GARCÍA ULL, «Deepfakes: El próximo reto en la detección de noticias falsas», *Anàlisi: Quaderns de Comunicació i Cultura*, núm. 24, 2021, p. 105 y KWEILIN T. Lucas, «Deepfakes and domestic violence: perpetrating intimate partner abuse using video technology», *Victims & Offenders*, vol. 17, núm. 5, 2022, p. 650.

[6] TOLOSANA ET AL., «Deepfakes and beyond: A survey of face manipulation and fake detection», *Information Fusion*, núm. 64, 2020, pp. 132-133 y HANCOCK, Jeffrey T. y BAILENSON Jeremy N., «The social impact of deepfakes», *Cyberpsychology, behavior, and social networking*, vol. 24, núm. 3, 2021, p. 151.

[7] KWOK/KOH, «Deepfake: a social construction of technology perspective», *Current Issues in Tourism*, vol. 24., núm. 3, 2021, pp. 1799-1800.

[8] Un ejemplo práctico de su aplicación puede rescatarse de la esfera de la publicidad donde el empleo de *deepfakes* puede favorecer la personalización de contenidos publicitarios audiovisuales dinámicos para cada persona consumidora. Además, en el caso del marketing, cabe la creación de probadores virtuales donde las personas compradoras podrían probarse virtualmente prendas en su propio cuerpo en lugar de en una figura modelo. Cuando se dificulta la producción de contenido en un lugar físico, como pudo ocurrir durante la pandemia, con el uso de *deepfakes* las agencias de publicidad pueden generar nuevos anuncios alterando secuencias grabadas con anterioridad o creando nuevas en formato hiperrealista desde un estudio. En el ámbito artístico, este tipo de aplicación de IA tiene un gran potencia si se asume la posibilidad de ocupar la ausencia de un artista apostando por su presencia a través de una representación artificial como expresión de creatividad y como oportunidad para favorecer la interacción y participación del público con la obra. KIETZMANN/MILLS/PLANGGER, «Deepfakes: perspectives on the future «reality» of advertising and branding», *International Journal of Advertising*, vol. 40, núm. 3, 2021, pp. 477-478 y MIHAILOVA, «To Dally with Dalí: Deepfake (Inter)faces in the Art Museum», *Convergence: The International Journal of Research into New Media Technologies*, vol. 27, núm. 4, 2021, pp. 885 y 895.

derechos fundamentales y contra la estabilidad institucional de los Estados afectando a cuestiones de política internacional y doméstica. Entre otros, puede señalarse la pornografía de venganza, el acoso, el chantaje, la propaganda, la manipulación del mercado o de procesos electorales, la difamación, la alteración de la opinión pública, los fraudes, los ataques a la seguridad nacional o las noticias falsas[9]. Además de apuntar hacia los usos de estas representaciones sintéticas, resulta conveniente matizar su ámbito subjetivo. Esto es, definir quiénes son los sujetos activos que los utilizan y quiénes los pasivos sobre los que se ejerce la manipulación.

GARCÍA ULL elabora una clasificación atendiendo a la autoría de las *deepfakes*. Señala a comunidades de aficionados (con fines de comedia y sátira o pornografía), actores políticos (en relación a posibles actos de competencia partidista, hackeos o desinformación), delincuentes y estafadores (por la creación perfiles falsos o la comisión de delitos financieros) y, lo que denomina como actores legítimos (en referencia a productores audiovisuales, creadores artísticos o agencias de publicidad), como los principales intervinientes en la producción de contenido falso[10].

La diversidad que se encuentra en los potenciales usuarios de *deepfakes*, no puede advertirse en quienes las padecen. Pese a los supuestos más conocidos en los que se vieron involucrados los expresidentes de Estados Unidos Barack Obama y Donald Trump o el CEO de Facebook, Mark Zuckerberg y otros más recientes que afectan al Papa Francisco o al presiente de Rusia Vladímir Putin y al presidente chino, Xi Jinping, lo cierto es que, según estudios recientes, el grupo objetivo de los primeros registros de *deepfakes* lo constituyen las mujeres, siendo el 96% de las *deepfakes* online de contenido sexual sin previo consentimiento y participando en el 99% de los casos mujeres[11]. El impacto de su uso está vinculado al acceso a la creación de contenido falso hiperrealista. La proliferación de grandes modelos generativos con una interfaz de alto nivel, es decir, que ofrecen una experiencia de usuario sencilla y directa como Dall-E, Midjourney o Stable Diffusion y aplicaciones como FakeApp, FaceSwap, DeepFaceLab posibilita que personas sin conocimientos especializados puedan crear contenido ficticio para el propósito que consideren[12], pudiendo constituir el ejercicio de violencia hacia las mujeres uno de ellos.

Resulta complejo dimensionar el impacto que puede llevar aparejado un fenómeno en plena evolución, como es el caso de las aplicaciones que emplean IA. Su valoración con datos específicos requiere de cierta toma de distancia del objeto de estudio para su problematización e identificación de las variables de análisis. La empresa holandesa Deeptrace, dedicada a la detección y supervisión de materiales creados artificialmente, destaca el rápido desarrollo de *deepfakes* tanto en términos de sofisticación tecnológica como de impacto social. Según sus cálculos, el número total de vídeos deepfake en línea ha experimentado un aumento de casi el 100% desde diciembre de 2017 a diciembre de 2018, alcanzando la cifra de 14.678. Como se ha avanzado, el empleo de estas técnicas para generar contenido pornográfico es mayoritario. Pese

[9] KWEILIN, «Deepfakes and domestic violence: perpetrating intimate partner abuse using video technology», *op. cit.*, p. 649; KWOK/KOH, «Deepfake: a social construction of technology perspective», *op. cit.*, pp. 1799-1800 y GARCÍA ULL, Francisco José, «Deepfakes: El próximo reto en la detección de noticias falsas», *op. cit.*, p. 106.

[10] GARCÍA ULL, «Deepfakes: El próximo reto en la detección de noticias falsas», *op. cit.*, p. 109.

[11] KWEILIN, «Deepfakes and domestic violence: perpetrating intimate partner abuse using video technology», *op. cit.*, p. 649.

[12] GÓMEZ DE ÁGREDA/FEIJÓO GONZÁLEZ/SALAZAR GARCÍA, «Una nueva taxonomía del uso de la imagen en la conformación interesada del relato digital. Deep fakes e inteligencia artificial», *El Profesional de la información*, vol. 30, núm. 2, 2021, p. 14.

a que el primer sitio web de pornografía deepfake se registró en febrero de 2018, el número total de visionados de representaciones audiovisuales de este tipo rebasa los 134 millones[13].

Informes de la Oficina Europea de Policía (Europol) y del Federal Bureau of Investigation (FBI) del año 2021 muestran su preocupación por el aumento del uso de contenido sintético en los próximos años a través de operaciones cibernéticas con una finalidad de manipulación sociopolítica y delictiva. Se prevé la incorporación de la IA en técnicas ya existentes para ampliar el alcance y escalar los ciberataques[14]. De hecho, los efectos de su empleo ya no son un futurible. En diciembre de 2022, se detuvo en la ciudad de Valladolid al primer pedófilo que generó a través de *deepfakes* «material de abuso sexual infantil de extrema dureza»[15].

1.2. Violencia contra las mujeres en la era digital

La violencia hacia las mujeres se transforma y se adapta atendiendo al contexto de su ejercicio, actualmente mediado por el uso de las tecnologías de la información y la comunicación (TIC). Con el cambio de paradigma generado por la irrupción de la IA surgen nuevos modos de violencia contra las mujeres con idéntica finalidad controladora. Como ocurre con la manipulación de imágenes, el fenómeno de la violencia contra las mujeres no es nuevo, lo es el procedimiento a través del cual se procura su ejercicio. La generalización del uso de las TIC se presenta como un nuevo medio mediante el cual es posible violentar a las mujeres y vulnerar sus derechos fundamentales. Según datos de Naciones Unidas presentes en un informe de 2015, cuando la tecnología aún no había conseguido el grado de desarrollo actual, un 73% de mujeres ya había sufrido algún tipo de violencia online. Incluye, además, que en el caso de la Unión Europea, el 18% de las mujeres ha sufrido una forma grave de violencia en Internet desde los 15 años, lo que equivale a unos 9 millones de mujeres[16]. En una guía para prevenir y responder ante este tipo de violencia, Naciones Unidas puntualiza que, aunque esta modalidad de ataques puede afectar a cualquier persona, las mujeres y las niñas son más vulnerables y experimentan la vivencia de verse violentadas de un modo más traumático cuyos efectos negativos pueden afectar a su estado conductual, emocional, mental, físico y social[17].

No solo las redes sociales se convierten en espacios virtuales para el ejercicio de la violencia sino que los juegos online exponen a las mujeres a situaciones y comportamientos abiertamente misóginos, violentos y atentarios contra sus derechos[18]. Así se recoge en la noticia publicada por el periódico *The Business Standard* donde se desvela el contenido de una página de juegos de Bangladesh en Facebook cuyo guión de los vídeos para el juego Grand Theft Auto V (GTA) reproduce expresiones como «¡Eh, chica inmoral, para! ¿Qué lleva puesto? ¡Te voy a dar una

[13] AJDER/PATRINI/CAVALLI/CULLEN, *The State of Deepfakes: Landscape, Threats, and Impact*, September 2019, p. 1.

[14] GOMES-GONÇALVES, «Los *deepfakes* como una nueva forma de desinformación corporativa – una revisión de la literatura», *IROCAMM*, vol. 5, núm. 2, 2022, p. 32.

[15] NAVARRO, «Detenido un pedófilo que usaba inteligencia artificial para crear material de abuso sexual infantil», *El País*, Diciembre 21, 2022. https://elpais.com/sociedad/2022-12-21/detenido-un-pederasta-que-usaba-inteligencia-artificial-para-crear-material-de-abuso-sexual-infantil.html

[16] UN BROADBAND COMMISSION FOR DIGITAL DEVELOPMENT WORKING GROUP ON BROADBAND AND GENDER, *Cyber violence against women and girls. A world-wide wake-up call*, UN Women, UNDP and ITU, 2015, pp. 15-16.

[17] DIALLO, *A Guide for Women and Girls to Prevent and Respond to Cyberviolence*, UN Women, November 2021, p. 4.

[18] DÍEZ GUTIÉRREZ, «Video games and gender-based violence», *Procedia-Social and Behavioral Sciences*, núm. 132, 2014, p. 59.

paliza!, mientras se golpea a las mujeres virtuales en bikini con un bate de béisbol»[19]. De hecho, la sección española de Amnistía Internacional ya en el año 2004 denunció el fomento de la violencia explícita y gratuita de dicho videojuego[20].

El Instituto Europeo de la Igualdad de Género (EIGE, por sus siglas en inglés) califica la ciberviolencia como una forma de violencia de género e incluye dentro de sus modalidades el ciberhostigamiento consistente en «compartir fotografías o vídeos íntimos de la víctima a través de internet o del teléfono móvil» de forma reiterada y la ciberexplotación, venganza pornográfica o pornografía no consentida que «implica la distribución en línea de fotografías o vídeos sexualmente explícitos sin el consentimiento de la persona que aparece en las imágenes»[21].

La Delegación del Gobierno para la Violencia de Género publicó en 2014 un incipiente estudio sobre *El ciberacoso como forma de ejercer la violencia de género en la juventud* para conocer el estado de la cuestión a través de la realización de grupos de discusión y entrevistas a jóvenes de entre 18 y 29 años. Según el informe, el ciberacoso en el marco de las relaciones de pareja es una práctica asentada que «supone una dominación sobre la víctima mediante estrategias humillantes que afectan a la privacidad e intimidad, además del daño que supone a su imagen pública», centradas principalmente en el chantaje emocional, las amenazas, los insultos, el deterioro de su imagen social y la posibilidad de forzar encuentros «casuales» a partir de su localización (con el consecuente riesgo de sufrir violencia offline)[22]. La facilitación de contenido íntimo no surge, según la información recabada, del desconocimiento de los potenciales usos perjudiciales que pueden hacerse a través de las redes, sino de la confianza, como gesto de amor[23], lo cual revela el modo en que se traspasa la frontera de la íntima confianza, abusando de ella, para provocar un daño.

En el núcleo de las *deepfakes* se encuentra el engaño, la intencionalidad de faltar a la verdad a sabiendas de la falsedad[24]. De este modo, podría entenderse el engaño como una fórmula de manipulación, como una manifestación más de violencia psicológica que puede afectar a derechos como el honor, la imagen, la integridad moral y que, consta, puede llevar a mujeres a quitarse la vida como ocurrió en el caso de la trabajadora de IVECO tras la difusión de vídeos íntimos en un grupo de WhatsApp de trabajo, asunto que quedó archivado al no poder identificar a la primera persona que los divulgó, ni existir pruebas de extorsión a la víctima[25]. Para la

[19] BILLAH, «Killing women for fun: How some Facebook gamers are inciting violence against women», *The Business Standard*, August 4, 2022. https://www.tbsnews.net/features/panorama/killing-women-fun-how-some-facebook-gamers-are-inciting-violence-against-women

[20] AMNISTÍA INTERNACIONAL, *Con la violencia hacia las mujeres no se juega. Videojuegos, discriminación y violencia contra las mujeres*, Sección Española, 2004, pp. 23-24.

https://www.amnistiacatalunya.org/edu/pdf/videojocs/04/vid-04-12.pdf

[21] INSTITUTO EUROPEO DE LA IGUALDAD DE GÉNERO, *La ciberviolencia contra mujeres y niñas*, 2017, pp. 2-3.

[22] DELEGACIÓN DEL GOBIERNO PARA LA VIOLENCIA DE GÉNERO, *El ciberacoso como forma de ejercer la violencia de género en la juventud*, Ministerio de Sanidad, Política Social e Igualdad. Centro de Publicaciones, 2014, pp. 4, 187, 191-192.

[23] Ibídem, p. 187.

[24] HANCOCK/BAILENSON, «The social impact of deepfakes», *op. cit.*, p. 149

[25] DOIAGA MONDRAGON ET AL., «Image-based abuse: Debate and reflections on the «Iveco Case» in Spain on Twitter», *Journal of interpersonal violence*, vol. 37(9-10), p. 7190.

Delegación de Gobierno, el sexting[26], como modalidad de ciberacoso contra las mujeres «es especialmente significativa y dañina puesto que, dada la forma viral de transmisión de la información en el mundo digital, en un breve lapso de tiempo se expande vertiginosamente y la audiencia supera el finito ámbito de amigos y conocidos»[27]. Señala BLANCO RUIZ, la prevalencia de la violencia psicológica a través de las redes sociales y cómo la digitalización de las situaciones violentas, intimidatorias o de control las convierte en actitudes más sutiles, menos evidentes, al tiempo que potencia que devengan conductas presentes las 24 horas del día debido a la omnipresencia de las pantallas[28].

Con todo, puede convenirse que la violencia contra las mujeres evoluciona según los avances en la digitalización de las relaciones y, en particular, atendiendo a los progresos y posibilidades que ofrecen los sistemas de IA. Por ello, resulta conveniente analizar los supuestos problemáticos que pueden derivarse de la unión entre *deepfake* y violencia de género partiendo de la hipótesis de investigación de que las *deepfakes* pueden convertirse en un facilitador de la violencia.

2. *Deepfakes* y violencia de género: tres categorías problemáticas

Poder profundizar en el binomio *deepfake*-violencia de género requiere puntualizar los principales factores que enmarcan esta nueva conexión. El primero de ellos, por su carácter estructural, lo conforma el sistema cisheteropatriarcal, que sitúa a las mujeres como cuerpos inanimados y lee su corporalidad como simple objeto de consumo[29]. Un consumo que, como especifican WAGNER y BLEWER, se materializa visualmente permitiendo eludir cualquier posibilidad de consentimiento o agencia por parte del rostro (y los cuerpos) que aparecen en las imágenes alteradas[30]. Un segundo elemento lo constituye el estado, denominado por WESTERLUND como *infoapocalipsis*, por el que no es posible distinguir lo que es real de lo que no lo es[31]. El desarrollo de imágenes artificiales hiperrealistas puede conducir a consolidar ese estado en el que sea altamente complejo diferenciar el producto ficticio del genuino, en el que la identidad y conducta de una mujer pueda ser manipulada con cierta facilidad e impunidad derivada de la dificultad de identificar la falsedad y de tener que combatir la incertidumbre y la confusión que estas producciones sintéticas generan. Por último, cabe mencionar los aspectos ya anotados relativos a la operatividad de las *deepfakes* relacionados con el engaño y la accesibilidad. Es importante destacar la motivación de engaño que envuelve esta técnica de creación de vídeos a partir de sistemas de IA, ya sea con una finalidad más o menos inocua, pero en el núcleo se halla la generación de contenido falso. Una posibilidad al alcance de cada vez más personas, a través

[26] Por sexing se debe entender como «el envío de fotos o videos de contenido erótico o carga sexual de manera consentida», una práctica que permanecería en el ámbito de las relaciones privadas sin relevancia penal, pero que se torna problemática cuando falla la presencia del consentimiento en la toma o difusión de las imágenes. 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: revista de derecho penal, procesal y penitenciario*, núm. 105, 2013, p. 28.

[27] DELEGACIÓN DEL GOBIERNO PARA LA VIOLENCIA DE GÉNERO, El ciberacoso como forma de ejercer la violencia de género en la juventud, op. cit., p. 188.

[28] BLANCO RUIZ, «Implicaciones del uso de las redes sociales en el aumento de la violencia de género en adolescentes», *Comunicación y medios*, núm. 30, 2014, pp. 128 y 135.

[29] SENENT/BUESO, «The banality of (automated) evil: critical reflections on the concept of forbidden knowledge in machine learning research», *Recerca. Revista de Pensament i Anàlis*i, vol. 27, núm, 2, 2022, p. 7.

[30] WAGNER/BLEWER, «The Word Real Is No Longer Real»: Deepfakes, Gender, and the Challenges of AI-Altered Video», *Open Information Science*, vol. 3, núm, 1, 2019, p. 33.

[31] WESTERLUND, «The emergence of deepfake technology: A review», *Technology Innovation Management Review*, vol. 9, núm. 11, p. 40.

de las plataformas de acceso libre, lo cual supone la socialización del engaño abriendo su ventana de producción al conjunto de la sociedad.

2.1. Desprestigio

Tal y como describen GÓMEZ DE ÁGREDA, FEIJÓO GONZÁLEZ y SALAZAR GARCÍA, la información de dudosa fiabilidad incide intensamente en la conformación crítica de la opinión pública sobre los distintos sucesos que acontecen, ya que pese a que este tipo de desinformación procedente de grupos de interés político-económicos no supera la tasa del 20% del total de contenidos que se difunden, sí alcanza unas cuotas de viralización próximas al 70% de las interacciones totales[32].

El estado actual del ecosistema informativo, en cierto modo preso de las noticias *clickbait* y empañado por la proliferación de *fake news*, sumado al plus de veracidad que conlleva acompañar una información con un documento gráfico visual[33], puede ser aprovechado por hombres maltratadores para generar *deepfakes* con la finalidad no exclusiva de la pornografía de venganza sino también para controlar, intimidar, aislar, avergonzar y microgestionar a las víctimas[34]. Este ejercicio de la violencia puede ocurrir en el seno de la pareja, pero también fuera de ella. Un ejemplo de desprestigio fue el intento de silenciamiento con la difusión de un vídeo pornográfico falso de la periodista de investigación Rana Ayyub al momento de denunciar públicamente las circunstancias políticas que rodeaban la violencia sexual a una niña de 8 años natural de Cachemira[35]. La estrategia del silencio puede ser utilizada también como instrumento de obstaculización para la presentación de denuncias por supuestos de violencia de género. No es descabellado pensar que los hombres maltratadores puedan generar *deepfakes* para poner en duda la versión de las futuras denunciantes («Tengo la prueba de que hubo consentimiento») y forzar el desistimiento («¿Quién te va a creer si eres una buscona?»)[36].

La afectación al prestigio puede revisarse atendiendo a un componente de género y a un elemento que podría subsumirse en una cuestión de clase. Apuntan CERDÁN MARTÍNEZ y PADILLA CASTILLO que mientras «[e]llas protagonizan falsas escenas eróticas y pornográficas; ellos, discursos y circunstancias relacionados con el humor o con la política, apareciendo normalmente vestidos. Ellas asoman en espacios privados e íntimos; ellos, en espacios públicos, ostentando el poder o un protagonismo sano. Ellas son cosificadas y sus rostros se pegan al cuerpo de una actriz despersonificada. Ellos tienen otro cuerpo, u otra voz, pero no pierden su esencia personal ni son tratados como objetos porque lo llamativo es lo que dicen o hacen. Ellas son sujetos pasivos; ellos son protagonistas activos y mueven la acción»[37]. Estas exhibiciones devuelven a las mujeres a un estadio inicial de la representación patriarcal más basta: el descrédito relacionado con la sexualidad por la incapacidad consustancial a la naturaleza femenina de alcanzar esferas públicas más elevadas como la comedia o la política y el retorno al enclaustramiento privado de cuatro

[32] GÓMEZ DE ÁGREDA/FEIJÓO GONZÁLEZ/SALAZAR GARCÍA, «Una nueva taxonomía del uso de la imagen en la conformación interesada del relato digital. Deep fakes e inteligencia artificial», *op. cit.*, pp. 6 y 14.

[33] Ibídem, pp. 2-3.

[34] KWEILIN, «Deepfakes and domestic violence: perpetrating intimate partner abuse using video technology», *op. cit.*, p. 648.

[35] MADDOCKS, «'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political deep fakes», *Porn Studies*, vol. 7, núm. 4, 2020, p. 415.

[36] PFEFFERKORN, «Deepfakes" in the Courtroom», *BU Pub. Int. LJ*, vol. 29, p. 255.

[37] CERDÁN MARTÍNEZ/PADILLA CASTILLO, «Historia del" fake" audiovisual:" deepfake" y la mujer en un imaginario falsificado y perverso», *Historia y comunicación social*, vol. 24, núm. 2, 2019, pp. 516-517.

paredes. Las mujeres que acceden a puestos de poder con cuotas de visibilización y representatividad son diana de técnicas sexistas que buscan ridiculizarlas o agredirlas en mayor medida que a los hombres, como ocurrió en el caso de la política estadounidense Nancy Pelosi cuyas declaraciones en actos públicos fueron manipuladas simulando un estado de embriaguez[38].

El factor de clase podría apreciarse en la diferente afectación de la difusión de *deepfakes* entre mujeres con popularidad y mujeres que viven una vida al margen de las cámaras, los micros y la fama. Las mujeres con renombre público cuentan con una presunción de buena reputación que las hace parcialmente inmunes a los contenidos mayoritariamente pornográficos que se destapan. Su público objetivo no asume ese tipo de información como verídica de forma que el perjuicio a su prestigio queda minimizado[39]. Además, cuentan con los medios de comunicación y las redes sociales como altavoces para desmentir y recuperar un relato que ha pretendido ser manipulado. Sin embargo, las mujeres, en general, no poseen ninguno de estos recursos: no hay un principio de *in dubio pro mulier* que pueda ofrecerles cierto margen para deconstruir el mensaje que se quiere difundir ni tampoco disponen de la oportunidad de defenderse en el espacio público, aumentando su vulnerabilidad y teniendo dificultades para obtener reparación[40].

2.2. Suplantación de identidad

La polivalencia de las *deepfakes* en cuanto a sus fines ha alcanzado la esfera de la ciberseguridad alertando sobre posibles problemas derivados de suplantar la identidad, por ejemplo, de la dirección ejecutiva de una empresa con el objetivo de operar con cierto tipo de actividades financieras, de personas realizando conductas fraudulentas o delictivas como forma de extorsión, también en el espacio de la pornografía no consentida[41].

De este modo, con la expansión de su uso, las prácticas de ciberacoso, descritas por PÉREZ MARTÍNEZ y ORTIGOSA BLANCH, podrían articularse a través de *deepfakes*. Consisten en falsificaciones fraudulentas de la identidad realizadas por hombres maltratadores tanto respecto a su propia imagen como a partir del reemplazo de la identidad de las mujeres. El primer supuesto lo constituye el engaño a las víctimas haciéndose pasar por amistades o personas conocidas para concertar un encuentro digital y llevar a cabo algún tipo de acoso online[42]. Este caso puede ser ampliado, considerando la posibilidad de que el encuentro pueda ser presencial, favoreciendo la comisión de alguna de las modalidades de violencia ya sea física o psicológica a través de coacciones o amenazas.

En cuanto a la suplantación de la identidad de las mujeres, las autoras describen la creación de perfiles o espacios falsos en los que la víctima comparte videos íntimos y realiza demandas y

[38] HARWELL, «Faked Pelosi videos, slowed to make her appear drunk, spread across social media», The *Washington Post*, May 24, 2019. https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/

[39] RAYMOND HARRIS, «Video on demand: what deepfakes do and how they harm», *Synthese*, núm. 199, 2021, p. 13386.

[40] MIOSOTIS SOTO, «Justice for Women: Deep fakes and Revenge Porn», *3rd Global Conference on Women's Studies*, Rotterdam: The Netherlands, 25-27 February 2022, p. 199.

[41] GARCÍA ULL, «Deepfakes: El próximo reto en la detección de noticias falsas», *op. cit.*, p. 111.

[42] PÉREZ MARTÍNEZ/ORTIGOSA BLANCH, «Una aproximación al ciberbullying», en GARCÍA GONZÁLEZ, Javier (coord.), *Ciberacoso: la tutela penal de la intimidad, la integridad y la libertad sexual en internet*, Tirant lo Blanch, 2010, p. 19.

ofertas sexuales explícitas[43]. Fruto de los nuevos desarrollos de las GAN, ya no es necesario contar con un repositorio de vídeos ni limitar la oferta y demanda a textos escritos o audios distorsionados. Con la aplicación *DeepNude* es posible «desnudar» a una mujer. A partir de un retrato y de una base de datos de 10.000 fotos de mujeres desnudas tomadas de internet, añade un cuerpo desnudo al rostro retratado[44]. La creación artificial de vídeos hiperrealistas multiplica las posibilidades de perpetuación de estas conductas: una técnica tecnológica depurada permite la depuración de las técnicas de violencia generando nuevos espacios de posibilidad del maltrato.

2.3. Simulación de situaciones ficticias constitutivas de delito

Manteniendo la misma lógica de distinción del sujeto activo manipulado por *deepfakes*, es posible anticipar otros supuestos ya catalogados como constitutivos de ciberacoso que podrían ser desarrollados en contra de las mujeres. En concreto, se trata de la creación de conductas tipificadas como delito en el Código Penal que involucran a víctima y maltratador.

Podría pensarse que, ante el conocimiento de la interposición de una denuncia por parte de una mujer, un investigado podría sofisticar la conocida «contradenuncia» a partir de la creación de un vídeo en el que aparece de forma gráfica la mujer cometiendo un delito contra su persona. Esta estrategia procesal de la defensa, en ocasiones carente de soporte probatorio concluyente, que se reduce a conflictos de pareja mal resueltos con porcentajes de sentencias con conformidad elevados[45], contaría esta vez con un material con el que se podría proceder a la imputación de un delito de agresiones mutuas. Para ORTUBAY FUENTES, la táctica de las denuncias cruzadas supone una instrumentalización del sistema penal, aprovechando los vacíos o imprecisiones de la Ley Orgánica 1/2004, de 28 de diciembre, de Medidas de Protección Integral contra la Violencia de Género (LOVG), y un resurgimiento de la tradición patriarcal que otorga al hombre, como cabeza de familia, la potestad de imponer su voluntad al resto de miembros por todos los medios[46].

Asimismo, como precisan FRANKS y WALDMAN, más allá de los vídeos de contenido sexual, no sería difícil imaginar representaciones en las que personas de minorías raciales o del colectivo LGBTIQ+ aparecieran cometiendo delitos como mecanismo de confirmación de sesgos y bulos construidos en torno a ellas[47]. Podría ser el caso de los supuestos de agresiones sexuales por parte de migrantes a mujeres españolas o la comisión de abusos sexuales a menores por parte de hombres homosexuales. Con este mismo patrón, podrían circular vídeos sobre comportamientos penalmente reprochables a las víctimas persiguiendo reacciones adversas y represalias jurídico-sociales contra ellas[48], relativas no solo a la posible iniciación de un proceso penal sino a la búsqueda del aislamiento de la víctima de su círculo de cercano familiar, de amistades, de trabajo como tipología de maltrato[49] y método de generación de vínculos de dependencia y subordinación más fuertes hacia el hombre maltratador.

---

[43] Ibídem.

[44] BAÑUELOS CAPISTRÁN, «Deepfake: la imagen en tiempos de la posverdad», *Revista Panamericana de Comunicación*, núm. 1, 2020, p. 55.

[45] ORTUBAY FUENTES, «Cuando la Respuesta Penal a la Violencia Sexista se Vuelve contra las Mujeres: las Contradenuncias», *Oñati Socio-legal series*, vol. 5, núm. 2, 2015, pp. 652 y 657.

[46] Ibídem, p. 665.

[47] FRANKS/WALDMAN, «Sex, lies, and videotape: Deep fakes and free speech delusions», *Maryland Law Review*, vol. 78, núm. 4, 2018, p. 896.

[48] PÉREZ MARTÍNEZ/ORTIGOSA BLANCH, «Una aproximación al ciberbullying», *op. cit*., p. 19.

[49] EXPÓSITO/MOYA, «Violencia de género», *Mente y cerebro*, núm. 48, 2011, p. 25.

3. Modelos anticipatorios: la articulación de una triple respuesta preventiva

La incidencia de las *deepfakes* en el terreno de la violencia contra las mujeres, obliga a anticipar un modelo de respuesta integral para prevenir el despliegue de una nueva modalidad de violencia de género que se beneficie de los avances en sistemas de IA Generativa. La propuesta es trifásica por dos razones fundamentales. Las *deepfakes* pueden entenderse como fenómeno tecnológico, pero también cultural, legal y ético[50], de modo que una única vía de actuación resulta en un abordaje parcial e incompleto que impide conjugar todos los elementos que componen esta nueva realidad[51]. El segundo motivo por el cual se apuesta por un enfoque multidisciplinar reside en la ineficacia contrastada del privilegio del Derecho, en particular del derecho penal, para erradicar la violencia contras las mujeres. Mientras que la LOVG se elaboró desde una perspectiva holística apelando a la educación, la sanidad y los medios de comunicación, además de a los tribunales como agentes vehiculares de la prevención, detección y eliminación de la violencia contra las mujeres, ha sido el sistema penal el que ha sufrido una hipertrofia a partir de las modificaciones del Código Penal y de la instauración de la denuncia como la mejor y, en ocasiones, única salida a la violencia. Pese al recorrido temporal de la ley, no se ha consolidado la necesaria coordinación interinstitucional ni los recursos materiales que son elementales para transitar una situación de violencia (renta básica, alternativas habitacionales, servicios de asistencia psicológica, acompañamiento jurídico…) y alcanzar su reparación.

Ambas justificaciones fuerzan a proponer un sistema de triple respuesta desde el ámbito: 1) jurídico, con especial atención a los problemas de prueba y de tipificación penal; 2) político, reflexionando sobre la incertidumbre informativa en la etapa de la posverdad, y 3) técnico, vinculado con la capacidad de diseñar otros sistemas GAN de detección de *deepfakes*.

3.1. Respuesta jurídica

El incremento en la sofisticación de la tecnología conlleva problemas de identificación al dificultar la detección de la manipulación «a ojo desnudo», como indican GÓMEZ DE ÁGREDA, FEIJOÓ GONZÁLEZ y SALAZAR GARCÍA[52]. Los índices de verosimilitud son tan elevados que no solo son indistinguibles las recreaciones de las figuras reales, sino que las imágenes creadas con IA tienden a generar mayor confianza[53]. Esta cuestión, aparentemente técnica, sobre la complejidad para localizar vídeos falsos se introduce de pleno en el ámbito judicial, en la fase dedicada a valorar el acervo probatorio propuesto por las partes y su autenticidad: ¿Tiene el conjunto de operadores jurídicos las herramientas necesarias para diferenciar un vídeo falso de uno verdadero? ¿Entienden qué son las *deepfakes* y la IA Generativa? Con el potencial uso generalizado de esta tecnología y las aplicaciones prácticas ya expuestas en los supuestos de violencia de género, es posible que en los juzgados se introduzca una prueba generada a partir de GAN o se pretenda la inadmisión de una propuesta alegando su manipulación[54]: ¿Dispone la Administración de Justicia de los medios suficientes para su tratamiento?

[50] BAÑUELOS CAPISTRÁN, «Deepfake: la imagen en tiempos de la posverdad», *op. cit.*, p. 52.

[51] Se habla de realidad porque las *deepfakes* van a estrechar, si no diluir, la línea entre realidad y ficción, entre verdad y falsedad, teniendo que impulsar nuevos instrumentos que disminuyan el grado de confusión e incertidumbre.

[52] GÓMEZ DE ÁGREDA/FEIJÓO GONZÁLEZ/SALAZAR GARCÍA, «Una nueva taxonomía del uso de la imagen en la conformación interesada del relato digital. Deep fakes e inteligencia artificial», *op. cit.*, p.15.

[53] FRANGANILLO, «Contenido generado por inteligencia artificial: oportunidades y amenazas», *op. cit.*, p. 7.

[54] PFEFFERKORN, «Deepfakes" in the Courtroom», *op. cit.*, p. 255.

Estos interrogantes ponen la atención en la especialización desde dos esferas complementarias. En primer lugar, se advierte la necesidad de un peritaje judicial avanzado, sofisticado e hiperexperto, requiriendo incluso de la propia IA, como se verá a continuación, para detectar los vídeos falsos. Es urgente poner a disposición de la judicatura personas expertas en investigación digital forense con conocimientos específicos en redes generativas adversariales que puedan ofrecer información clara para conformar el criterio de quienes ostentan la potestad jurisdiccional para aceptar o denegar una prueba o, en su caso, dotarle del valor probatorio correspondiente. En este sentido, advierte PFEFFERKORN sobre las consecuencias para el proceso de la autentificación de vídeos apuntando hacia la prolongación del litigio y el aumento de los gastos asociados a la realización de diligencias adicionales por parte de personas expertas[55].

En segundo lugar, resulta oportuno profundizar en la alfabetización digital de sus señorías. Tal y como dispone el artículo 301.3 de la Ley Orgánica 6/1985, de 1 de julio, del Poder Judicial (LOPJ) para ingresar en la Carrera Judicial por la categoría de juez es preciso superar una oposición libre y un curso teórico. Este último viene regulado en el artículo 307 LOPJ y está estructurado en tres fases consecutivas: un primer programa de formación multidisciplinar teórico-práctico de nueve meses, unas prácticas tuteladas durante seis meses y medio y, por último, un segundo período de prácticas, esta vez desempeñando funciones de sustitución y refuerzo con una duración de cuatro meses. Interesa en este punto la capacitación que debe garantizar el Consejo General del Poder Judicial (CGPJ) a través de un Plan Docente de Formación Inicial impartido por la Escuela Judicial (art. 433 bis LOPJ).

En la programación formativa correspondiente a la 72ª Promoción de la Carrera Judicial para el curso 2022-2023 no se oferta un contenido específico respecto a los sistemas de IA y su incorporación al plano legal de forma teórica y práctica (pese a tratarse de una cuestión ampliamente debatida por la doctrina). En el Plan Docente de Formación Inicial se apuesta por la formación en competencias sustentada en la triple alianza «saber», «saber hacer» y «saber ser» o, en otros términos, conocimientos, habilidades y actitudes. En la esfera de las habilidades sí se encuentra, no obstante, una referencia a las TIC, al asumir por parte de la Escuela Judicial que «el desarrollo digital exige la adquisición de habilidades específicas para trabajar en ese entorno y con dichas herramientas»[56]. Asimismo, en el listado de actividades transversales de carácter multidisciplinar hay una dedicada a la sociedad de la información. Presenta un módulo introductorio «a los conceptos básicos del lenguaje informático y a los aspectos más destacables de la nueva realidad digital» entre los que podrían incluirse las GAN y las *deepfakes* y, desde una vertiente procesal, se analiza «cuáles son los medios de obtención de pruebas relacionados con el uso de las nuevas tecnologías, poniendo especial énfasis en la forma en que las partes han obtenido la prueba, con especial referencia a la fiabilidad y a la licitud de la misma»[57], un requisito ya mencionado que precisa ser estudiado con atención.

En cuanto a los tipos penales que permitirían la persecución de estas prácticas, la conducta penalmente reprochable más próxima se asemeja a las detalladas en el artículo 197 del Código Penal (CP), delito de descubrimiento de secreto (apartado 1) o de sexting ajeno o difusión no

[55] Ibídem, p. 275.

[56] CONSEJO GENERAL DEL PODER JUDICIAL. Escuela Judicial, Plan Docente de Formación Inicial. 72.ª Promoción de la Carrera Judicial: curso 2022-2023, p. 26.

[57] Ibídem, pp. 94-95.

consentida de imágenes íntimas (apartado 7) en función de si las imágenes se han obtenido con o sin permiso. En el primer caso, se castigaría con penas de prisión de 1 a 4 años de prisión y multa de 12 a 24 meses a quien, sin permiso, se apoderase de imágenes o vídeos íntimos de una persona. La pena sería de 2 a 5 años si se procediera a su difusión (art. 197.3 CP) y se impondría en su mitad superior cuando los hechos descritos en los apartados anteriores afecten a datos de carácter personal que revelen la vida sexual de la víctima (197.5 CP). En el supuesto de sexting ajeno, la obtención de las imágenes o grabaciones audiovisuales sí cuenta con la autorización de las personas involucradas, pero no su difusión, revelación o cesión a terceros por lo que cuando la divulgación menoscabe gravemente la intimidad personal de la persona, se impondrá la pena de 3 meses a 1 año o multa de 6 a 12 meses (art. 197.7 CP), fijando el cómputo en su mitad superior cuando los hechos hubieran sido cometidos en el seno de una relación de pareja o análoga en los términos establecidos por la LOVG. Con la entrada en vigor de la Ley Orgánica 10/2022, de 6 de septiembre, de garantía integral de la libertad sexual (LOGILS) se introduce un párrafo en el apartado 7 del artículo 197 por el cual se castiga con pena de multa de 1 a 3 meses a quienes reciban esas imágenes o vídeos y procedan a difundirlas.

La principal diferencia entre los delitos descritos y el caso de las *deepfakes* es que en este último no se trata de la obtención de imágenes íntimas[58] (con o sin consentimiento y, en el caso de existir anuencia, no es la propia víctima la que produce el contenido y lo envía ni es ella la que consiente que otra persona capte el contenido audiovisual[59]) sino que se trata de su creación artificial a partir de imágenes públicas o privadas. En consecuencia, no parece que el supuesto de las *deepfakes* pueda ser subsumible en los tipos previstos en el artículo 197 CP, pero sí puede ser considerado como una subcategoría (pudiendo añadirse un nuevo apartado al artículo referente al modo de generación del contenido audiovisual) que requiere de la articulación de un marco legal para prevenir su desarrollo y, en caso de ocurrir, penalizar la conducta.

Tal y como defiende LLORIA GARCÍA, «los cambios son necesarios sobre todo en algunos ámbitos, puesto que 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»[60]. En este sentido mantiene la autora la necesidad de reinterpretar el bien jurídico «intimidad» a la luz de los avances que han procurado la transformación de un escenario absolutamente analógico a uno absolutamente tecnológico, de modo que han de preverse las oportunidades de lesión más grave que genera el entorno virtual, la facilidad para hacer llegar -o generar- el contenido íntimo por el uso de la tecnología[61] y la potencial afectación a múltiples bienes jurídicos: «libertad (amenazas y coacciones), a la intimidad (revelación de secretos, apropiación de contraseñas), al honor (difusión de imágenes y comentarios), la integridad moral (degradación continuada y

[58] Para una precisión acerca de lo requisitos de la intimidad de las imágenes consultar: LLORIA GARCÍA, «La difusión de imágenes íntimas sin consentimiento en derecho penal español», en ASOCIACIÓN ARGENTINA DE PROFESORES DE DERECHO PENAL (coord.), *Derecho Penal y Pandemia. Homenaje al Prof. Julio B. Maier*, Ediar, 2020, pp. 213 ss.

[59] VALENZUELA GARCÍA, «El rol de las TIC en el delito de "sexting" problemas de aplicabilidad del artículo 197.7 del Código Penal», en ARÁNGUEZ SÁNCHEZ, Tasia y OLARIU, Ozana (coord.), *Feminismo digital: violencia contra las mujeres y brecha sexista en Internet*, Dykinson, 2021, p. 450.

[60] LLORIA GARCÍA, «Delitos y redes sociales: los nuevos atentados a la intimidad, el honor y la integridad moral (especial referencia al «sexting»)», *op. cit.*, p. 25.

[61] LLORIA GARCÍA, «La difusión de imágenes íntimas sin consentimiento en derecho penal español», *op. cit.*, p. 211.

permanente, acoso) [...] o la usurpación de personalidad (presentación del sujeto ante la red con la utilización de una identidad correspondiente a un tercero)[62]».

3.2. Respuesta política

El desfase temporal entre los ritmos acelerados que gestionan la vida política y la dedicación tecno-jurídica pausada que supone atribuir la autoría de un montaje o constatar una manipulación[63], motiva la necesaria anticipación preventiva frente a la consolidación de un contexto de escepticismo generalizado y de constante duda que, más allá de los macroriesgos advertidos en un paradigma securitario[64] que «ha aceptado, deseado, necesitado y asumido el peligro como medio natural»[65], debilita la cohesión social y la democracia[66]. Para ejemplificar esta afectación sociopolítica de los desórdenes informativos se utilizan, como casos paradigmáticos, las elecciones presidenciales en Estados Unidos y el referéndum del Brexit en Reino Unido en 2016, también como muestra de las prácticas de engaño imperantes en la era de la posverdad[67].

El fenómeno de la posverdad, cuya conceptualización resulta controvertida al aproximarse al concepto mismo de verdad y mentira[68], no es reciente. Como estrategia destinada a «establecer una idea como verdadera a partir de una manipulación de información, hechos, actos, emociones, actores y escenarios mediáticos»[69], ha estado presente en la historia de la humanidad desde los años sesenta. Sin embargo, puede hallarse la novedad en el escenario que potencia el despliegue de estas técnicas de engaño. En una época en la que se producen y difunden ingentes cantidades de datos y se enarbolan como si fueran emanaciones de la verdad[70], una época en la que se ha hecho de la cultura de compartir (fotos, estados, vídeos, ubicaciones...) un imperativo categórico[71], una época en la que el empleo de sistemas de IA potencia la diseminación de contenidos generados artificialmente desconociendo su autoría[72]. En este entorno es en el que se

[62] LLORIA GARCÍA, «Delitos y redes sociales: los nuevos atentados a la intimidad, el honor y la integridad moral (especial referencia al «sexting»)», *op. cit.*, pp. 25 y 27.

[63] GÓMEZ DE ÁGREDA/FEIJÓO GONZÁLEZ/SALAZAR GARCÍA, «Una nueva taxonomía del uso de la imagen en la conformación interesada del relato digital. Deep fakes e inteligencia artificial», *op. cit.*, p. 10.

[64] Tales como la utilización de *deepfakes* como ataques de otros países, de grupos terroristas, de partidos en la oposición, de grandes corporaciones financieras a un Estado motivados por estrategias geopolíticas, de desestabilización, por intereses económicos, por competencia electoral, por presión lobista que influyen en la soberanía nacional.

[65] QUINTERO OLIVARES, «Los delitos de riesgo en la politica criminal de nuestro tiempo», en ARROYO ZAPATERO/NEUMANN/NIETO MARTIN (coord.), *Crítica y justificación del Derecho Penal en el cambio de Siglo*, Universidad de Castilla-La Mancha, 2003, p. 241.

[66] GONZÁLEZ ARENCIBIA/HERNÁNDEZ VELÁZQUEZ, «Una mirada crítica al pensamiento de la postverdad», *Serie Científica de la Universidad de las Ciencias Informáticas*, vol. 14, núm. 7, 2021, pp. 18-19.

[67] RUBIO NÚÑEZ, «Los efectos de la posverdad en la democracia», *Revista de Derecho Político*, núm. 103, 2018, p. 193.

[68] CARRERA, «Estratagemas de la posverdad», *Revista Latina de Comunicación Social*, núm. 73, 2018, pp. 1469 ss.

[69] BAÑUELOS CAPISTRÁN, «Deepfake: la imagen en tiempos de la posverdad», *op. cit.*, p. 53.

[70] RODRÍGUEZ FERRÁNDIZ, «Posverdad y fake news en comunicación política: breve genealogía», *Profesional de la información*, vol. 28, núm. 3, 2019, p. 9.

[71] Ibídem.

[72] GONZÁLEZ ARENCIBIA/HERNÁNDEZ VELÁZQUEZ, «Una mirada crítica al pensamiento de la postverdad», *op. cit.,* p. 22.

insertan las *deepfakes* erosionando la capacidad del público para discernir la verdad de la falsedad[73].

Los postulados de la posverdad se encargan de ensombrecer la correspondencia entre enunciados y hechos y apelan a la emocionalidad y a las creencias personales para la conformación de la opinión pública, alejando de ella los hechos objetivos[74]. Aunque pueda estar desdibujándose la popular falacia «si no lo veo, no lo creo» por el uso generalizado de *deepfakes* no es menos cierto que esta técnica de creación de vídeos ha aprovechado tanto la necesidad de corroboración gráfica como la interpelación a las emociones para potenciar su impacto. Tal y como expone DEL FRESNO GARCÍA, se «necesita de un mecanismo de legitimación en el que se persigue naturalizar una epistemología basada en las emociones políticas, dado que las emociones y sentimientos son reales los hechos que los provocan, los desórdenes informativos, tienen que ser reales»[75].

La sobreexposición a una enorme cantidad de información de la que es complicado distinguir la veracidad de la falsedad y de la manipulación intencionada, moldea la percepción del mundo y la comprensión de la realidad[76]. No solo tiene un efecto sobre la inteligibilidad del entorno, sino que la falta de confianza en las noticias que se reciben (y, por tanto, en los medios de comunicación que dotan de contenido al derecho constitucional a la información) puede afianzar el recurso a los sesgos de confirmación[77], así como el descarte de imágenes genuinas como falsas por el simple hecho de considerar que aquello que no se quiere creer debe ser falso[78].

La repercusión de las *deepfakes* en la cosmovisión y construcción de una realidad social compartida, sitúa en el centro de la agenda política la necesidad de dotar a la ciudadanía de las herramientas para ser capaz de detectar la mentira y la falsedad (si es que es humanamente posible). Requiere también de los poderes públicos la promulgación de una cultura fuerte del manejo de la información y una sana sospecha crítica sobre las fuentes de información que generen una alerta sobre el origen de los datos y la necesidad de contrastarlos, en especial, cuando la difusión de información ficticia vulnera derechos fundamentales como en el caso estudiado de la violencia contra las mujeres.

En la LOGILS, impulsada por el Ministerio de Igualdad, el artículo 7 dedicado a la prevención y sensibilización en el ámbito educativo, establece en su apartado segundo que en todas las etapas educativas no universitarias se «incluirán contenidos formativos sobre el uso adecuado y crítico de internet y las nuevas tecnologías, destinados a la sensibilización y prevención de las violencias sexuales, la protección de la privacidad y los delitos cometidos a través de las nuevas tecnologías de la información y la comunicación promoviendo una educación en la ciudadanía digital mediante la consecución de competencias digitales adaptadas a nivel correspondiente del tramo de edad». El apartado tercero extiende esta formación en títulos universitarios oficiales cuando resulte coherente conforme a las competencias inherentes a los mismos.

---

[73] FRANKS/WALDMAN, «Sex, lies, and videotape: Deep fakes and free speech delusions», *op. cit.*, p. 893.

[74] Esta definición se contiene en el Diccionario Oxford el cual declaró «posverdad» como palabra del año en 2016. Conceptualización disponible en: https://languages.oup.com/word-of-the-year/2016/

[75] DEL FRESNO GARCÍA, «Desórdenes informativos: sobreexpuestos e infrainformados en la era de la posverdad», *El profesional de la información (EPI)*, vol. 28, núm. 3, 2019, p. 3.

[76] Ibídem, p. 8.

[77] Ibídem, p. 7.

[78] GARCÍA ULL, «Deepfakes: El próximo reto en la detección de noticias falsas», *op. cit.*, p. 111.

Esta previsión normativa ya supone un avance puesto que el impacto de *deepfakes* podría quedar perfectamente contemplado para ponerlo en conocimiento de las generaciones más jóvenes, nativas digitales y potenciales usuarias finales de estas aplicaciones. También, desde una perspectiva de género, ya que cabe la concienciación sobre la utilización inconsentida de imágenes de mujeres para crear contenido sexual, lo cual coligue con el apartado primero de dicho precepto en el que se fija la inclusión en el sistema educativo español de «contenidos basados en la pedagogía feminista sobre educación sexual e igualdad de género y educación afectivo-sexual para el alumnado, apropiados en función de la edad». Como en toda materia que requiere de una transformación sustancial, la base educativa deviene el estrato básico sobre el que iniciar los cambios colectivos, para lo cual los medios de comunicación cumplen una función esencial como promotores y altavoces de campañas de concienciación y sensibilización dirigidas a la sociedad.

Conviene precisar que la comunicación y la información son las bases para una sociedad democrática. Siguiendo las reivindicaciones de RUBIO NÚÑEZ se concluye que: 1) «[l]a sociedad es esencialmente comunicación, hasta el punto de que sin comunicación no hay sociedad», ya que sobre ella se construyen todo tipo de relaciones (laborales, económicas, industriales, culturales, religiosas, de ocio, personales que configuran la sociedad)[79]; 2) «[u]n acuerdo sobre la existencia de la verdad y la posibilidad de alcanzarla vuelve a ser el fundamento indispensable de una verdadera democracia», por lo que es imprescindible reducir el impacto de las estrategias de desinformación redefinidas por los progresos en IA que han propiciado un clima de descrédito generalizado, de pérdida de referencias informativas válidas y de sentimentalización de las decisiones políticas[80].

3.3. Respuesta técnica

Los problemas para detectar la manipulación de la información audiovisual que se está reproduciendo devuelven la responsabilidad de su identificación al espacio de las GAN. La alarma sobre la proliferación de *deepfakes* ha promovido iniciativas de *deep-checking*. Una de ellas consiste en incrustar una huella digital permanente que advierta de que el material ha sido alterado con IA[81]. También es posible encontrar inconsistencias entre los movimientos de los labios y el discurso de audio o en las variaciones de las imágenes[82], incluso en el parpadeo de los ojos. Asimismo, el análisis forense de imágenes puede utilizarse para rastrear la historia de un documento audiovisucal acudiendo a extremos como el formato de almacenamiento, el proceso de adquisición o cualquier otro de postprocesado que pueda haber dejado un rastro único de los datos[83]. Aunque los resultados son todavía limitados y no generalizables, el proceso de

[79] RUBIO NÚÑEZ, «Los efectos de la posverdad en la democracia», *op. cit.*, p. 195.

[80] Ibídem, p. 227.

[81] FRANGANILLO, «Contenido generado por inteligencia artificial: oportunidades y amenazas», *op. cit.*, p. 11 y GÓMEZ DE ÁGREDAFEIJÓO GONZÁLEZ/SALAZAR GARCÍA, «Una nueva taxonomía del uso de la imagen en la conformación interesada del relato digital. Deep fakes e inteligencia artificial», *op. cit.*, p. 4.

[82] KORSHUNOV/SÉBASTIEN, «Deepfakes: a new threat to face recognition? Assessment and detection», *op. cit.*, p. 4.

[83] ALBAHAR/ALMALKIPP, «Deepfakes: threats and countermeasures systematic review», *Journal of Theoretical and Applied Information Technology*, vol. 97, núm. 22, 2019, pp. 3247-3248.

individualización podría realizarse utilizando las mismas redes generativas adversariales que se emplean para la generación de *deepkafes*[84].

La detección técnica de *deepfakes* parece ser una de las acciones más prometedoras a corto y medio plazo, también porque su implementación y consecución son más inmediatas que una alteración del orden jurídico o político, si se comparan con los tiempos parlamentarios, las negociaciones de grupos políticos o la efectiva concienciación e (in)formación de la población.

En todo caso, estos avances en ingeniería computacional podrían contribuir muy positivamente a superar los obstáculos anticipados en los dos ámbitos anteriores, forzando la idea inicial de interdependencia entre las tres esferas de actuación y enfatizando la multidisciplinariedad como perspectiva esencial para el abordaje de problemáticas poliédricas. Podrían ser puestos al servicio de la Administración de Justicia ante las dificultades probatorias ya reseñadas, así como a disposición de quienes se encargan del diseño e implementación de políticas públicas, ya que podrían hacer uso de ellos los medios de comunicación de titularidad pública como filtro para la difusión de información veraz, pero también impulsar leyes y códigos de buenas prácticas tanto para medios de comunicación convencionales como para redes sociales y plataformas de entretenimiento que obligaran a su correcta distinción.

4. Conclusiones: ¿ver para creer?

La Inteligencia Artificial Generativa, en su modalidad de *deepfakes* en particular, supone una alteración del sistema informativo, ya que obliga a retirar el cargo de certificación de la realidad que históricamente han ostentado los documentos audiovisuales[85]. Con la creación ficticia de contenido, ¿será funcional *ver para creer*?

A raíz de identificar los potenciales usos negativos que se pueden idear ante la aparición de nuevos dispositivos tecnológicos, es imprescindible adoptar enfoques prospectivos desde la multidisciplinariedad que permitan anticipar los riesgos que pueden afectar a los derechos de la ciudadanía, especialmente de aquellos colectivos menos dignificados, y que pueden tensionar los valores fundamentales de las democracias europeas.

Para FRANKS y WALDMAN, «[n]o hacer nada contra toda expresión dañina en la era digital dista mucho de ser una no intervención liberal; más bien, es una opción normativa que perpetúa el poder de las mayorías atrincheradas contra las minorías vulnerables»[86]. En el caso de la violencia contra las mujeres, es posible constatar cómo la tecnología *deepfake* permite extender, y retroalimentar, la violencia desde un nodo offline a otro online. La violencia muta y encuentra otros medios de ejercicio de control por lo que, advertida su metamorfosis, corresponde a los poderes públicos adoptar las medidas político-jurídicas adecuadas que prevengan su materialización. Sin necesidad se activar un estado de alarma, pero asumiendo el impacto por la introducción de *deepfakes* en el proceso judicial, tanto desde un punto de vista sustancial como

[84] REMYA REVI/VIDYA/WILSCY, «Detection of Deepfake Images Created Using Generative Adversarial Networks: A Review», en PALESI/TRAJKOVIC/JAYAKUMARI/JOSE (ed.), *Second International Conference on Networks and Advances in Computational Technologies*, Springer, 2021, p 33.

[85] BAÑUELOS CAPISTRÁN, «Deepfake: la imagen en tiempos de la posverdad», *op. cit.*, p. 54.

[86] Traducción propia. Texto disponible en: FRANKS/WALDMAN, «Sex, lies, and videotape: Deep fakes and free speech delusions», *op. cit.*, p. 893.

procedimental (con especial atención en la práctica de la prueba) cabe repensar la actuación del conjunto de los operadores jurídicos (abogacía, judicatura, fiscalía y peritos) que se enfrentan a la tarea de demostrar la autenticidad o falsedad de un vídeo[87].

En palabras de DEL FRESNO GARCÍA, «[l]a verdad fáctica es una empresa colectiva y su vigencia y extensión demuestran el éxito de esa empresa en nuestra historia como especie»[88]. El estado de duda permanente, que no se asemeja a la sana crítica, sino que deriva en desconfianza hacia los recursos informativos y que agrava sesgos de confirmación se constituye como un escenario amenazador para los sistemas democráticos que beben de la cohesión y los pactos sociales. En consecuencia, es preciso encontrar mecanismos (jurídicos, políticos y tecnológicos) que despejen la incertidumbre y retornen percepciones sólidas y racionales a la ciudadanía.

La complejidad que entrañan los avances tecnológicos reclama una acción conjunta desde diferentes áreas de conocimiento y por parte de distintos actores públicos. Por ello, la triada propuesta pretende conformar una visión integral problematizando un supuesto de hecho concreto a partir de la identificación de tres usos controvertidos de *deepfakes* contra las mujeres -desprestigio, suplantación de identidad y simulación de delitos- y formulando alternativas de solución desde la vigencia de los derechos humanos reconocidos a las mujeres. La erradicación de la violencia contra las mujeres pasa, necesariamente, por la identificación de nuevos riesgos, el análisis de sus efectos y la articulación de alternativas que prevengan y garanticen su bienestar.

5. Bibliografía

AJDER, Henry, PATRINI, Giorgio, CAVALLI, Francesco y CULLEN, Laurence*, The State of Deepfakes: Landscape, Threats, and Impact*, September 2019.

ALBAHAR, Marwan y ALMALKIPP, Jameel, «Deepfakes: threats and countermeasures systematic review», *Journal of Theoretical and Applied Information Technology*, vol. 97, núm. 22, 2019, pp. 3242-3250.

AMNISTÍA INTERNACIONAL, *Con la violencia hacia las mujeres no se juega. Videojuegos, discriminación y violencia contra las mujeres*, Sección Española, 2004.
https://www.amnistiacatalunya.org/edu/pdf/videojocs/04/vid-04-12.pdf

BARONA VILAR, Silvia, «Inteligencia Artificial o la algoritmización de la vida y de la justicia: ¿solución o problema?», *Rev. Boliv. de Derecho*, núm. 28, 2019, pp. 18 ss.

BAÑUELOS CAPISTRÁN, Jacob, «Deepfake: la imagen en tiempos de la posverdad», *Revista Panamericana de Comunicación*, núm. 1, 2020, pp. 51 ss.

BILLAH, Masum, «Killing women for fun: How some Facebook gamers are inciting violence against women», *The Business Standard*, August 4, 2022.

[87] PFEFFERKORN, «Deepfakes" in the Courtroom», *op. cit.*, pp. 254, 258 y 267.

[88] DEL FRESNO GARCÍA, «Desórdenes informativos: sobreexpuestos e infrainformados en la era de la posverdad», *op. cit.*, p. 9.

https://www.tbsnews.net/features/panorama/killing-women-fun-how-some-facebook-gamers-are-inciting-violence-against-women

BLANCO RUIZ, María Ángeles, «Implicaciones del uso de las redes sociales en el aumento de la violencia de género en adolescentes», *Comunicación y medios*, núm. 30, 2014, pp. 124 ss.

CARRERA, Pilar, «Estratagemas de la posverdad», *Revista Latina de Comunicación Social*, núm. 73, 2018, pp. 1469 ss.

CERDÁN MARTÍNEZ, Víctor y PADILLA CASTILLO, Graciela, «Historia del" fake" audiovisual:" deepfake" y la mujer en un imaginario falsificado y perverso», *Historia y comunicación social*, vol. 24, núm. 2, 2019, pp. 505 ss.

CONSEJO GENERAL DEL PODER JUDICIAL. Escuela Judicial, *Plan Docente de Formación Inicial. 72.ª Promoción de la Carrera Judicial: curso 2022-2023*, pp. 1 ss. Disponible en: https://www.poderjudicial.es/cgpj/es/Temas/Escuela-Judicial/Formacion-Inicial/La-fase-presencial/Plandocente-de-formacion-inicial-72--Promocion-Carrera-Judicial--curso-2022-2023

DELEGACIÓN DEL GOBIERNO PARA LA VIOLENCIA DE GÉNERO, *El ciberacoso como forma de ejercer la violencia de género en la juventud*, Ministerio de Sanidad, Política Social e Igualdad. Centro de Publicaciones, 2014, pp. 1 ss.

DEL FRESNO GARCÍA, Miguel, «Desórdenes informativos: sobreexpuestos e infrainformados en la era de la posverdad», *El profesional de la información (EPI)*, vol. 28, núm. 3, 2019, pp. 1 ss.

DIALLO, Amira, *A Guide for Women and Girls to Prevent and Respond to Cyberviolence*, UN Women, November 2021.

DÍEZ GUTIÉRREZ, Enrique Javier, «Video games and gender-based violence», *Procedia-Social and Behavioral Sciences*, núm. 132, 2014, pp. 58 ss.

DOIAGA MONDRAGON, Nahia, DOSIL SANTAMARIA, Maria, BELASKO TXERTUDI, Maitane, & ALONSO SAEZ, Israel, «Image-based abuse: Debate and reflections on the «Iveco Case» in Spain on Twitter», *Journal of interpersonal violence*, vol. 37(9-10), pp. 7178 ss.

EXPÓSITO, Francisca y MOYA, Miguel, «Violencia de género», *Mente y cerebro*, núm. 48, 2011, pp. 20 ss.

FRANGANILLO, Jorge, «Contenido generado por inteligencia artificial: oportunidades y amenazas», *Anuario ThinkEPI* 16, 2022, pp. 1 ss.

FRANKS, Mary Anne y WALDMAN, Ari Ezra, «Sex, lies, and videotape: Deep fakes and free speech delusions», *Maryland Law Review*, vol. 78, núm. 4, 2018, pp. 892 ss.

GARCÍA ULL, Francisco José, «Deepfakes: El próximo reto en la detección de noticias falsas», *Anàlisi: Quaderns de Comunicació i Cultura*, núm. 24, 2021, pp. 103 ss.

GÓMEZ DE ÁGREDA, Ángel, FEIJÓO GONZÁLEZ, Claudio Antonio y SALAZAR GARCÍA, Idoia Ana, «Una nueva taxonomía del uso de la imagen en la conformación interesada del relato digital. Deep fakes e inteligencia artificial», *El Profesional de la información*, vol. 30, núm. 2, 2021, pp. 1 ss.

GOMES-GONÇALVES, Sónia, «Los *deepfakes* como una nueva forma de desinformación corporativa – una revisión de la literatura», *IROCAMM*, vol. 5, núm. 2, 2022, pp. 22-38.

GONZÁLEZ ARENCIBIA, Mario y HERNÁNDEZ VELÁZQUEZ, Miguel Ramón, «Una mirada crítica al pensamiento de la postverdad», *Serie Científica de la Universidad de las Ciencias Informáticas*, vol. 14, núm. 7, 2021, pp. 17 ss.

GOODFELLOW, Ian, et al., «Generative adversarial networks», *Communications of the ACM*, vol. 63, núm. 11, 2020, pp. 139 ss.

HANCOCK, Jeffrey T. y BAILENSON Jeremy N., «The social impact of deepfakes», *Cyberpsychology, behavior, and social networking*, vol. 24, núm. 3, 2021, pp. 149 ss.

HARWELL, Drew, «Faked Pelosi videos, slowed to make her appear drunk, spread across social media», The *Washington Post*, May 24, 2019.
https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/

INSTITUTO EUROPEO DE LA IGUALDAD DE GÉNERO, *La ciberviolencia contra mujeres y niñas*, 2017, pp. 1 ss.

KIETZMANN, Jan, MILLS, Adam J. y PLANGGER Kirk, «Deepfakes: perspectives on the future «reality» of advertising and branding», *International Journal of Advertising*, vol. 40, núm. 3, 2021, pp. 473 ss.

KORSHUNOV, Pavel y SÉBASTIEN Marcel, «Deepfakes: a new threat to face recognition? Assessment and detection», 2018, pp. 1 ss. arXiv preprint arXiv:1812.08685

KWOK, Andrei y KOH, Sharon, «Deepfake: a social construction of technology perspective», *Current Issues in Tourism*, vol. 24., núm. 3, 2021, pp. 1798 ss.

KWEILIN, Lucas, «Deepfakes and domestic violence: perpetrating intimate partner abuse using video technology», *Victims & Offenders*, vol. 17, núm. 5, 2022, pp. 647 ss.

LLORIA GARCÍA, Paz, «Delitos y redes sociales: los nuevos atentados a la intimidad, el honor y la integridad moral (especial referencia al «sexting»)», *La ley penal: revista de derecho penal, procesal y penitenciario*, núm. 105, 2013, pp. 24 ss.

- «La difusión de imágenes íntimas sin consentimiento en derecho penal español», en ASOCIACIÓN ARGENTINA DE PROFESORES DE DERECHO PENAL (coord.), *Derecho Penal y Pandemia. Homenaje al Prof. Julio B. Maier*, Ediar, 2020, pp. 209 ss.

MADDOCKS, Sophie, «'A Deepfake Porn Plot Intended to Silence Me': exploring continuities between pornographic and 'political deep fakes», *Porn Studies*, vol. 7, núm. 4, 2020, pp. 415 ss.

MIHAILOVA, Mihaela, «To Dally with Dal´ı: Deepfake (Inter)faces in the Art Museum», *Convergence: The International Journal of Research into New Media Technologies*, vol. 27, núm. 4, 2021, pp. 882 ss.

MIOSOTIS SOTO, Santana, «Justice for Women: Deep fakes and Revenge Porn», *3rd Global Conference on Women's Studies*, Rotterdam: The Netherlands, 25-27 February 2022, pp. 113 ss. Disponible en: https://www.dpublication.com/wp-content/uploads/2022/02/27-10177.pdf

NAVARRO, Juan, «Detenido un pedófilo que usaba inteligencia artificial para crear material de abuso sexual infantil», *El País*, Diciembre 21, 2022. https://elpais.com/sociedad/2022-12-21/detenido-un-pederasta-que-usaba-inteligencia-artificial-para-crear-material-de-abuso-sexual-infantil.html

NIEVA FENOLL, Jordi, *Inteligencia artificial y proceso judicial*, Marcial Pons, Madrid, 2018.

ORTUBAY FUENTES, Miren, «Cuando la Respuesta Penal a la Violencia Sexista se Vuelve contra las Mujeres: las Contradenuncias», *Oñati Socio-legal series*, vol. 5, núm. 2, 2015, pp. 645 ss.

PÉREZ MARTÍNEZ, Ana y ORTIGOSA BLANCH, Reyes, «Una aproximación al ciberbullying», en GARCÍA GONZÁLEZ, Javier (coord.), *Ciberacoso: la tutela penal de la intimidad, la integridad y la libertad sexual en internet*, Tirant lo Blanch, 2010, pp. 13 ss.

PFEFFERKORN, Riana, «Deepfakes" in the Courtroom», *BU Pub. Int. LJ*, vol. 29, pp. 245 ss.

QUINTERO OLIVARES, Gonzalo, «Los delitos de riesgo en la politica criminal de nuestro tiempo», en ARROYO ZAPATERO, Luis, NEUMANN Ulfrid y NIETO MARTIN, Adan (coord.), *Crítica y justificación del Derecho Penal en el cambio de Siglo*, Universidad de Castilla-La Mancha, 2003, pp. 241 ss.

RAYMOND HARRIS, Keith, «Video on demand: what deepfakes do and how they harm», *Synthese*, núm. 199, 2021, pp. 13373 ss.

REMYA REVI, K., Vidya, K. R. y Wilscy, M., «Detection of Deepfake Images Created Using Generative Adversarial Networks: A Review», en PALESI, Maurizio, TRAJKOVIC, Ljiljana, JAYAKUMARI, J. y JOSE, John (ed.), *Second International Conference on Networks and Advances in Computational Technologies*, Springer, 2021, pp. 25 ss.

RODRÍGUEZ FERRÁNDIZ, Raúl, «Posverdad y fake news en comunicación política: breve genealogía», *Profesional de la información*, vol. 28, núm. 3, 2019, pp. 1 ss.

RUBIO NÚÑEZ, Rafael, «Los efectos de la posverdad en la democracia», *Revista de Derecho Político*, núm. 103, 2018, pp. 191 ss.

SENENT, Rosa y BUESO, Diego, «The banality of (automated) evil: critical reflections on the concept of forbidden knowledge in machine learning research», *Recerca. Revista de Pensament i Anàlis*i, vol. 27, núm, 2, 2022, pp. 1-26.

SURDEN, Harry, «Artificial intelligence and law: An overview», *Georgia State University Law Review*, núm. 35, 2019, pp. 1305 ss.

TOLOSANA, Ruben, et al., «Deepfakes and beyond: A survey of face manipulation and fake detection», *Information Fusion*, núm. 64, 2020, pp. 131 ss.

UN BROADBAND COMMISSION FOR DIGITAL DEVELOPMENT WORKING GROUP ON BROADBAND AND GENDER, *Cyber violence against women and girls. A world-wide wake-up call*, UN Women, UNDP and ITU, 2015.

VALENZUELA GARCÍA, Noelia, «El rol de las TIC en el delito de "sexting" problemas de aplicabilidad del artículo 197.7 del Código Penal», en Aránguez Sánchez, Tasia y Olariu, Ozana (coord.), *Feminismo digital: violencia contra las mujeres y brecha sexista en Internet*, Dykinson, 2021, pp. 440 ss.

WAGNER, Travis L. y BLEWER Ashley, «The Word Real Is No Longer Real»: Deepfakes, Gender, and the Challenges of AI-Altered Video», *Open Information Science*, vol. 3, núm, 1, 2019, pp. 32 ss.

WESTERLUND, Mika, «The emergence of deepfake technology: A review», *Technology Innovation Management Review*, vol. 9, núm. 11, pp. 39 ss.