Mhjournal
MIGUEL HERNÁNDEZ COMMUNICATION JOURNAL
ISSN: 1989-8681

Fecha recepción: 03/04/2024
Fecha revisión: 24/07/2024
Fecha de publicación: 30/07/2024

Deepfakes sexuales: impacto, prevención y perspectivas de género en el entorno digital

Cecilia Barba Arteaga | barbaarteagacecilia@gmail.com
Universitat Pompeu Fabra

Palabras clave
"Revisión bibliográfica" ; "Deepfakes sexuales"; "Inteligencia Artificial"; "Violencia sexual"; "Abuso Sexual Basado en Imágenes"

Sumario

Resumen

En los últimos años, ha habido un notable aumento en la creación y difusión de deepfakes, especialmente en el ámbito de contenido sexual en línea. Estos *deepfakes* han sido utilizados principalmente para crear contenido no consensuado de naturaleza sexual. Esta investigación examina la problemática de los *deepfakes* sexuales desde diversas perspectivas a partir de una revisión bibliográfica sistematizada, destacando su papel en la perpetuación de la violencia sexual y el sesgo de género al sexualizar y objetivar a las mujeres sin su consentimiento. Se discuten las implicaciones de la continuación de los abusos sexuales, la falta de consentimiento en el entorno digital, el impacto en las víctimas y el papel de esta tecnología en la perpetuación de la violencia sexual. Se describen distintas estrategias para frenar la creación y difusión de *deepfakes* sexuales, incluyendo métodos de detección, marcos legales y alfabetización mediática. Abordar esta problemática de manera integral y colaborativa es crucial para mitigar los impactos negativos de esta tecnología emergente y proteger los derechos de las víctimas.

Cómo citar este texto:

Cecilia Barba Arteaga (2024): Deepfakes sexuales: impacto, prevención y perspectivas de género en el entorno digital, en *Miguel Hernández Communication Journal, Vol. 15* (2), pp. 229 a 244. Universidad Miguel Hernández, UMH (Elche-Alicante).
DOI: 10.21134/zt4eht31

Sexual Deepfakes: Impact, Prevention and Gender Perspectives in the Digital Environment

Cecilia Barba Arteaga | barbaarteagacecilia@gmail.com

Universitat Pompeu Fabra

Keywords
"Literature Review"; "Sexual Deepfakes"; "Artificial Intelligence"; "Sexual Violence"; "Image-based Sexual Abuse"

Summary

Abstract

In recent years, there has been a notable increase in the creation and dissemination of deepfakes, especially in the area of online sexual content. These deepfakes have mainly been used to create non-consensual content of a sexual nature. This research examines the issue of sexual deepfakes from a variety of perspectives based on a systematised literature review, highlighting their role in perpetuating sexual violence and gender bias by sexualising and objectifying women without their consent. It discusses the implications of the continuation of sexual abuse, the lack of consent in the digital environment, the impact on victims and the role of this technology in perpetuating sexual violence. Strategies to curb the creation and dissemination of sexual deepfakes are described, including detection methods, legal frameworks and media literacy. Addressing this issue in a comprehensive and collaborative manner is crucial to mitigating the negative impacts of this emerging technology and protecting victims' rights.

How to cite this text:

Cecilia Barba Arteaga (2024): Deepfakes sexuales: impacto, prevención y perspectivas de género en el entorno digital, en *Miguel Hernández Communication Journal, Vol. 15* (2), pp. 229 a 244. Universidad Miguel Hernández, UMH (Elche-Alicante).
DOI: 10.21134/zt4eht31

1. Introducción y justificación

En los últimos seis años, hemos sido testigos de un incremento notable en la creación y circulación de *deepfakes*, tanto en su nivel tecnológico como en su cantidad (Rousay, 2023). La emergencia y propagación de *deepfakes* sexuales constituye un fenómeno perturbador y creciente en el ámbito de la pornografía en línea y la privacidad digital.

Los *deepfakes* son generados mediante el uso de tecnología de procesamiento de gráficos y técnicas avanzadas de aprendizaje profundo, como las redes neuronales recurrentes (RNN, por sus siglas en inglés) o las redes generativas adversarias (GAN, por sus siglas en inglés), con el fin de producir vídeos extremadamente realistas pero completamente falsos (Simón Soler, 2023). Sin embargo, incluso sin el uso de una tecnología avanzada, la cual no está al alcance de todos, ya existen formas de manipulación contextual que pueden ser replicadas fácilmente por cualquier persona (Paris & Donovan, 2019). El contexto inicial donde se observaron los *deepfakes* fue en el ámbito de la pornografía.

Si bien no todos los *deepfakes* son pornográficos, un estudio citado por Emily van der Nagel encontró que el "96% de las imágenes *deepfake* son pornografía no consensuada" (2020: 424). Se trata de una forma de representación de vídeo *deepfake* en donde se coloca el rostro de una persona, a menudo una celebridad femenina, aunque no siempre, sobre el cuerpo de otra con una intención sexual. Las personas representadas en los medios están involucradas en actos sexuales, los genitales están representados o los medios se producen o comparten con fines de gratificación sexual (Harper *et al.*, 2021). Es importante destacar que este desafío no se circunscribe únicamente a personalidades conocidas o figuras públicas, sino que también impacta a mujeres en situaciones cotidianas. El anonimato y la accesibilidad a la tecnología *deepfake* hacen que cualquier mujer pueda ser objeto de este tipo de abuso, lo que genera preocupaciones significativas en torno a la privacidad y la seguridad en línea para todas las mujeres (Gosse & Burkell, 2020; Rousay, 2023).

El primer caso documentado de contenido *deepfake amateur* surgió en línea en 2017, cuando un usuario de Reddit publicó varias imágenes manipuladas que superponían los rostros de celebridades femeninas en cuerpos de actrices de la industria pornográfica (Paris & Donovan, 2019). En 2020, un bot impulsado por inteligencia artificial se hizo disponible de forma gratuita en la plataforma de mensajería Telegram. Una revisión de los contenidos falsos creados y compartidos públicamente utilizando esta tecnología identificó más de 104.800 imágenes pornográficas falsas de diferentes mujeres en los seis meses posteriores a su lanzamiento (Flynn *et al.*, 2021).

Parte de la notoriedad generalizada de los *deepfakes* proviene precisamente de que mayoritariamente se desarrollan para crear vídeos sexuales no consensuados en línea. De hecho, se ha informado que la mayor parte de las personas afectadas y perjudicadas en la pornografía *deepfake* son mujeres (Flynn *et al.*, 2021), principalmente porque los algoritmos solo se han entrenado con sus imágenes (Jacobsen & Simpson, 2023). Estos *deepfakes* podrían representar una nueva forma de abuso sexual digital y convertirse, por tanto, en otro medio a través del cual las mujeres pierden el control sobre su imagen y son objeto de objetivación y sexualización (Jacobsen & Simpson, 2023).

El predominio de los *deepfakes* sexuales no consensuados y su falta de representación en los debates políticos y sociales refleja una tendencia histórica a ignorar o minimizar las preocupaciones de las mujeres en relación con la violencia sexual, lo que subraya la importancia de abordar las inequidades sistémicas de género en este contexto (Rousay, 2023). A pesar de la prevalencia y gravedad de los *deepfakes* sexuales, existe una notable falta de investigación académica sobre el tema y sus efectos perjudiciales asociados (Flynn *et al.*, 2022). Entre las investigaciones que se han realizado sobre esta problemática, se ha identificado una necesidad urgente de desarrollar estrategias de prevención y mitigación para abordar este tipo de abuso digital.

El presente artículo pretende explorar la problemática de los *deepfakes* sexuales desde una perspectiva de género, pues afecta mayoritariamente a mujeres, a partir de una revisión bibliográfica sistematizada. El objetivo principal ha sido revisar la situación de la producción académica sobre los *deepfakes* sexuales en aquellos estudios que trataran también la violencia sexual.

2. Metodología

El propósito principal ha sido examinar el estado actual de la producción académica relacionada con los *deepfakes* de contenido sexual y responder a las siguientes preguntas:

- ¿Son los *deepfakes* pornográficos considerados una nueva forma de violencia sexual que presenta un sesgo de género? ¿Cuáles son los precedentes de estos abusos y cuál es el rol del consentimiento en el entorno digital?

- ¿De qué manera afecta esta nueva forma de abuso sexual digital a las víctimas y a las motivaciones de los perpetradores?

- ¿Cuál es la respuesta que se está desarrollando para frenar la creación y difusión de *deepfakes* sexuales?

Para llevar a cabo esta revisión bibliográfica se tuvo en cuenta el marco de trabajo Framework Resiste-CHS, el cual se encuentra explicado en tres informes: "Revisiones bibliográficas sistematizadas en Ciencias Humanas y Sociales. 1: Fundamentos" (Codina, 2020a); "Revisiones bibliográficas sistematizadas en Ciencias Humanas y Sociales. 2: Búsqueda y Evaluación" (Codina, 2020b); "Revisiones sistematizadas en Ciencias Humanas y Sociales. 3: Análisis y Síntesis de la información cualitativa" (Codina, 2020c). El Framework Resiste-CHS es una guía estructurada para revisiones sistematizadas en ciencias sociales, enfocada en investigaciones cualitativas o conceptuales. Utiliza el Framework SALSA y se centra en cuatro fases: Búsqueda, para identificar y seleccionar documentos relevantes usando bases de datos y criterios específicos; Evaluación, para valorar la calidad metodológica de los estudios seleccionados; Análisis, para generar resúmenes sistemáticos de cada obra en forma de tablas o diagramas; y Síntesis, para crear un producto nuevo y valioso mediante la integración e interpretación de la información recopilada, aportando un alto valor añadido.

Es importante destacar que se han seguido los criterios del marco de trabajo Framework Resiste-CHS (Codina, 2020c), los cuales han sido utilizados para guiar la búsqueda de

artículos relevantes, establecer los criterios de selección y elegibilidad según requisitos predefinidos, y llevar a cabo la síntesis y análisis de los resultados obtenidos, pero también se han adaptados las líneas a esta investigación en particular.

Las plataformas seleccionadas para buscar producción científica en bases de datos académicas han sido Web of Science, Scopus y Google Scholar por ofrecer los resultados más completos y extensos. El marco de trabajo utilizado recomienda "utilizar siempre Scopus + Web of Science del grupo general" y "como precaución añadida, consultar Google Scholar para identificar trabajos académicos relevantes" (Codina, 2020b). Los criterios de búsqueda, tal como se describen aquí, fueron en inglés, ya que de esta forma aparecen textos tanto en inglés como en español. Debían aparecer en título del artículo, *abstract* y palabras clave:

- Búsqueda para "Violencia Sexual": ("sexual violence" OR "sexual abuse") AND ("gender" OR "sexual assault") AND ("online violence" OR "cyberbullying").

- Búsqueda para "Deepfakes": ("deepfakes" OR "false image generation" OR "media manipulation") AND ("technology" OR "multimedia content").

Al ser las preguntas de investigación muy amplias y al poderse responder desde distintas perspectivas, los criterios de análisis y síntesis debían abarcar toda investigación que tratara la violencia sexual y los *deepfakes* sexuales. Es por ello que no se ha sesgado por ámbito académico o metodología utilizados, únicamente debían estar publicados en inglés y castellano, y considerar la intersección ya nombrada. La búsqueda de la revisión bibliográfica se desarrolló durante el mes de diciembre del año 2023 en las bases de datos ya nombradas. Los resultados de búsqueda sin la selección posterior fueron: 25 documentos Scopus, con publicaciones desde el 2019 a 2023, teniendo este último año la mayor cantidad de artículos; 22 documentos en Web of Science, desde el 2018 a 2023, siendo 2021 el año con más publicaciones. En Google Scholar se seleccionaron un total de 10 documentos que cumplían los criterios de esta revisión.

Se seleccionaron un total de 18 publicaciones para el análisis y síntesis de esta revisión bibliográfica, descartando aquellos que no trataran los *deepfakes* sexuales y la violencia sexual, así como los artículos que estaban repetidos en las distintas bases de datos. Esto dejaba fuera, por ejemplo, investigaciones sobre el uso de la Inteligencia Artificial como forma de mitigar la violencia sexual en la prevención de crímenes, lo cual respondía de forma pertinente a la búsqueda realizada pero no entraba en la intersección entre violencia sexual y *deepfakes* sexuales. Luego, los criterios de selección, de forma más específica, han sido: tratar el uso de la Inteligencia Artificial en la creación de *deepfakes*; tratar la violencia sexual; cumplir con umbrales de calidad y rigor; estar publicados en inglés y castellano. Seguidamente se da el número de publicaciones por tipo de investigación: 6 revisiones jurídicas y legislativas; 6 revisiones de la literatura; 2 revisiones bibliográficas; 2 análisis de la cobertura mediática; 1 salud; 1 estudio sobre las víctimas.

Cabe señalar que dos de esas publicaciones no formaban parte de los resultados de búsqueda, sino que fueron añadidos más tarde por contener información significativa y valiosa para la investigación: "A systematic review of the current knowledge regarding revenge pornography and non-consensual sharing of sexually explicit media" (Walker & Sleath, 2017) y "Personality, Attitudinal, and Demographic Predictors of Non-consensual

Dissemination of Intimate Images" (Karasavva & Forth, 2021). Se consideró necesario incluir estos dos artículos aunque no formaran parte de la búsqueda inicial ya que, como se especificará más adelante, uno de los resultados de la búsqueda fue que los *deepfakes* sexuales entran dentro del Abuso Sexual Basado en Imágenes. Era entonces necesario completar la revisión con algún estudio que tratara este tipo de violencia.

3. Resultados

La primera respuesta que arroja la revisión bibliográfica es que los *deepfakes* pornográficos no consensuados son reconocidos como formas de violencia sexual. Específicamente, son clasificados dentro de la categoría del Abuso Sexual Basado en Imágenes (ASBI), que abarca la producción, difusión o amenaza de difusión de material sexual explícito sin el consentimiento de la persona afectada y puede manifestarse en diversos contextos (Jarvis Cooper, 2022). Esta modalidad de abuso digital ha sido identificada como una nueva manifestación de violencia de género, donde las mujeres son sexualizadas y convertidas en objeto sin consentimiento (Jacobsen & Simpson, 2023).

3.1. Violencia sexual: Abuso Sexual Basado en Imágenes

Esta forma de abuso sexual ha migrado al espacio digital, facilitando su perpetración y causando graves daños a las víctimas y abarcando diversas actividades, como la grabación no consensuada, la captura de imágenes sexuales sin permiso, la difusión no autorizada y la sextorsión. La introducción de la tecnología *deepfake* ha ampliado estas prácticas, posibilitando la creación y manipulación de datos para generar reproducciones digitales falsas con propósitos pornográficos (Okolie, 2023).

El ASBI, como concepto fundamental, representa una infracción grave contra la privacidad y la dignidad de las personas, generando impactos psicológicos inmediatos y repercusiones sociales y profesionales a largo plazo. Una de sus variantes que más investigación académica genera es la pornografía de venganza, donde se difunden fotografías íntimas tomadas supuestamente de manera consensuada, pero difundidas en línea sin el consentimiento de la víctima (Walker & Sleath, 2017; Jarvis Cooper, 2022).

3.1.1. Falta de consentimiento

Los *deepfakes* pornográficos se conectan estrechamente con la pornografía de venganza, compartiendo similitudes en cuanto a la falta de consentimiento, el impacto psicológico en las víctimas y la naturaleza sexual de las imágenes. Sin embargo, la aparición de la tecnología *deepfake* significa que ya no hay necesidad de que el perpetrador posea imágenes íntimas "reales" de su víctima. Los creadores de *deepfakes* solo necesitan imágenes suficientes del rostro de su objetivo, lo que hace que "cualquier mujer pueda aparecer en la pornografía" (Gosse & Burkell, 2020, p. 4).

Una de las bases que lleva a situar a los *deepfakes* como violencia sexual es la falta de consentimiento, que se comparte en todas las representaciones de ASBI. Precisamente por eso es importante apuntar que, a pesar de ser coloquialmente llamados "pornografía *deepfake*", hay estudios que abogan por el término "*deepfakes* sexuales", precisamente para enfatizar la naturaleza abusiva de dicho contenido, destacando la falta de consentimiento involucrado (Rousay, 2023).

Una forma de invalidar la falta de consentimiento es responsabilizando a las víctimas del abuso. La perspectiva de género destaca un fenómeno importante: la tendencia a responsabilizar a las víctimas, principalmente mujeres. "En las sociedades patriarcales, se insta a las mujeres a responsabilizarse personalmente de su propia seguridad frente a la violencia física y sexual" (Jacobsen & Simpson, 2023, p. 5). Algunos textos abordan este tema mediante el concepto de "discurso desviado" asociado al *sexting*, resaltando la importancia de cambiar el enfoque, dejando de culpar principalmente a las personas que comparten imágenes íntimas y dirigiéndose más bien hacia la problemática de la distribución no consensuada de dichas imágenes (Walker & Sleath, 2017; Karasavva & Forth, 2021).

3.1.2. Víctimas de estos abusos

Otro punto común son las víctimas de estos abusos. Es esencial, y así insisten los estudios, desmentir la idea errónea de que los *deepfakes* sexuales no consensuados son crímenes sin víctimas. La profunda huella del abuso de *deepfakes* incluye trauma psicológico, daño a la reputación e incluso suicidio. Las mujeres son tres veces más propensas a enfrentar distintas formas de abuso en línea que los hombres y más de dos veces más propensas a sufrir formas graves de abuso (Laffier & Rehman, 2023). Sobre los *deepfakes* sexuales, y de manera más específica, "el informe de DeepTrance muestra que el 99% de las víctimas son mujeres, aunque los vídeos deepfake no pornográficos (por ejemplo, publicados en YouTube) también están presentes en la red y afectan a hombres, difundiendo desinformación o ciberacoso" (Mania, 2022). Es decir, los *deepfakes* que no son de carácter sexual afectan tanto a hombres como a mujeres mientras que los *deepfakes* sexuales o pornográficos afectan casi íntegramente a mujeres.

En términos de impacto psicológico, las víctimas de *deepfakes* experimentan trastornos de estrés postraumático, depresión, ansiedad y otros problemas de salud mental (Lucas, 2022; Laffier & Rehman, 2023; Rousay, 2023). Se señala que el miedo constante a la propagación del contenido abusivo en línea puede llevar a las mujeres a vivir en un estado de ansiedad perpetua, afectando negativamente su calidad de vida y su capacidad para recuperarse del trauma (Lucas, 2022).

Se señala que las prácticas de pornografía de venganza exponen las desigualdades de género y la cosificación de las mujeres en los medios digitales, lo cual podemos transportar también a los *deepfakes* sexuales. La lucha contra este tipo de pornografía involuntaria representa un desafío para la sociedad de la información y destaca la necesidad de abordar estas cuestiones desde una perspectiva de género (Martínez Sánchez, 2023). El abuso cibernético refleja entendimientos culturales y sociales más amplios sobre género y estatus, con prácticas misóginas que insisten en la inferioridad de las mujeres (Laffier & Rehman, 2023).

Hay estudios que enfatizan que la naturaleza profundamente arraigada del ciberabuso basado en género y los *deepfakes* reflejan y refuerzan las normas sociales y culturales que subyacen a la desigualdad de género. Los *deepfakes*, en particular, reflejan y perpetúan la objetivación de las mujeres como objetos de deseo y gratificación sexual, socavando su agencia y autonomía (Laffier & Rehman, 2023) y convirtiéndose en una herramienta más para ejercer control y poder sobre las víctimas (Lucas, 2022).

Se discuten, por tanto, tres tipos de daños para las mujeres en estos estudios: angustia emocional y psicológica, pérdida de autonomía sobre el propio cuerpo y reputación, y la conexión entre los *deepfakes* y otros posibles delitos (Gosse & Burkell, 2020; Flynn *et al.*, 2021).

3.1.3. Motivaciones de los abusos

En los estudios que investigaban las motivaciones de los perpetradores del abuso sexual basado en imágenes, incluyendo la pornografía de venganza y los *deepfakes*, se describían distintas causas, que van desde el placer sexual hasta la venganza y la intimidación. Estas incluyen la búsqueda de satisfacción física o psicológica, el ejercicio de poder para causar daño emocional a la víctima, eludir la necesidad de consentimiento, buscar venganza después de una ruptura, demostrar masculinidad, realizar sextorsión y dañar la reputación social. Estas motivaciones resaltan la complejidad y gravedad del abuso sexual basado en imágenes, especialmente con la tecnología de *deepfake*, que dificulta el control y la eliminación del contenido falso (Okolie, 2023).

Otra de las motivaciones que es de considerable importancia destacar es el beneficio económico, la cual no comparte, en principio, con otros abusos sexuales en línea. Como ejemplo, es sonado e investigado el caso de la exposición accidental de Brandon "Atrioc" Ewing, un *streamer* de Twitch, de pornografía *deepfake* que había visto previamente de conocidas *streamers* de Twitch. Este incidente "puso de manifiesto la explotación de deepfakes de mujeres por parte de los creadores con fines lucrativos, pues Ewing había revelado inadvertidamente a su audiencia en directo el sitio de suscripción de pago por visión que alojaba dicho contenido" (Laffier & Rehman, 2023).

Se destaca que la generación de *deepfake* puede estar motivada no solo por el control financiero y físico, sino también por la curiosidad, la compulsividad sexual o un interés sexual específico. Se plantea que la generación de *deepfake* pornográfica podría ser utilizada para la gratificación sexual personal (Harper *et al.*, 2019).

3.2. Intensificación de los abusos en línea por la Inteligencia Artificial

Los estudios señalan que la amplificación en línea juega un papel crucial al evaluar la gravedad de una violación de la privacidad sexual en línea. Este tipo de contenido, el cual viola la privacidad sexual, puede ser fácilmente descubierto por usuarios a través de motores de búsqueda, y, una vez publicado en plataformas de redes sociales, tiende a ser compartido y reenviado múltiples veces. Los algoritmos de las plataformas sociales pueden propagar aún más el contenido no consensuado, y los procesos para eliminar dicho contenido tienden a ser lentos y poco efectivos. Además, incluso después de la eliminación, el contenido puede persistir en resultados de motores de búsqueda y otras plataformas, lo que hace que la violación de la privacidad en línea sea continua y duradera para la víctima (Jarvis Cooper, 2022; Martínez Sánchez, 2023).

Otra de las preocupaciones reflejadas en las investigaciones es que, aún sin tecnología avanzada, existen formas contextuales de manipulación fácilmente reproducidas por cualquiera, como el *software* de animación Adobe After Effects o proyectos de código

abierto como FakeApp, FaceSwap y DeepFace Lab. Estas herramientas permiten a cualquier persona generar vídeos falsos mediante el aprendizaje automático (Paris & Donovan, 2019). El acceso a la tecnología y el aumento de casos de la pornografía de venganza y los *deepfakes* sexuales se relaciona con la disponibilidad generalizada de programas informáticos y contenido pornográfico en línea, lo que facilita la creación y difusión de estos materiales (Mania, 2022).

Enfrentados a esta idea, existen otros estudios que, aunque reconocen la accesibilidad de estas manipulaciones, enfatizan en lo complejo que sigue siendo el proceso de creación de *deepfakes*. Jacquelyn Burkell y Chandell Gosse señalan que, aunque los avances en inteligencia artificial han simplificado y democratizado el proceso de crear *deepfakes*, lo cual plantea preocupaciones sobre el consentimiento y el control de la imagen de las personas, esta creación continúa siendo un proceso técnico muy complejo y deliberado (2019). A pesar de la existencia de proyectos de código abierto, se resalta que crear *deepfakes* requiere una considerable curva de aprendizaje y un cuidadoso procesamiento de imágenes. Este mismo artículo insiste en que el debate debería centrarse en la deliberación de la creación de *deepfakes* y no en su democratización:

> Aunque sostenemos que la tecnología no tiene toda la culpa, tampoco estamos diciendo que esté totalmente libre de culpa. La facilidad y accesibilidad para compartir información, e imágenes en particular, aumenta sin duda las consecuencias de los deepfakes. La cuestión es que, incluso con la ayuda del software disponible, la creación de deepfakes requiere trabajo, y es un trabajo profundamente intencionado: no ocurre accidentalmente ni de forma automática (Burkell & Gosse, 2019).

Para concluir con este apartado, conviene aclarar que el entorno cultural en el que emergen los *deepfakes* está impregnado de misoginia y la objetivación de las mujeres en la esfera digital y visual. Aunque la tecnología en sí misma no posee una inclinación misógina inherente, su empleo en la creación de *deepfakes* sexuales refleja la mirada masculina y machista subyacente:

> La misoginia no está "integrada" en la tecnología, sino que la decisión de utilizarla para crear deepfakes sexuales recae en los usuarios y refleja la cultura misógina en la que se despliega la tecnología (Burkell & Gosse, 2020, p. 2).

Si bien la tecnología de *deepfake* es novedosa, esta representa una continuación de la manera en que las imágenes de las mujeres han sido utilizadas para controlarlas y restringir su representación en la sociedad. Los *deepfakes* pueden ser interpretados como una forma de estilización algorítmica del cuerpo femenino, particularmente sexualizado, donde los rostros y cuerpos se vuelven cada vez más intercambiables. Con los *deepfakes*, los cuerpos y rostros se convierten en componentes primarios para la reconfiguración y estilización perpetua de lo que se describe como "mirada masculina" -término utilizado por la teórica Laura Mulvey (1989) en cine- algorítmica (Jacobsen & Simpson, 2023).

Aunque los *deepfakes* sexuales refuerzan la percepción de las mujeres como simples objetos sexuales, esta dinámica se enmarca en un contexto cultural y digital que, a lo largo de la historia, ha considerado a las mujeres como sujetos pasivos, para ser observados, en lugar de individuos con capacidad de observación propia. "Los deepfakes ocupan un espacio controvertido entre la ruptura y la continuidad, entre la objetivación femenina siempre emergente y la que ya ha surgido" (Jacobsen & Simpson, 2023, p. 12).

La utilización de deepfakes para generar contenido sexual no consensuado se encuentra intrínsecamente vinculada con esta cultura, así como con prácticas más amplias de abuso basado en imágenes (Burkell & Gosse, 2020).

3.3 *Respuesta política y técnica ante los deepfakes sexuales*

La respuesta política y técnica ante la proliferación de *deepfakes*, especialmente en el ámbito de la pornografía y el acoso sexual en línea, se ha centrado en varios frentes con el objetivo de abordar los desafíos planteados por esta tecnología disruptiva.

3.3.1 Marcos de detección

La creciente sofisticación de esta tecnología presenta desafíos en la verificación de la autenticidad de los medios (Laffier & Rehman, 2023), por lo que una de las principales respuestas propuestas en los estudios ha sido el desarrollo de métodos de detección de *deepfakes*. Se propone una solución técnica que se centre en el desarrollo de sistemas de inteligencia artificial capaces de detectar y prevenir la creación y difusión de *deepfakes*. Esto podría incluir el diseño de algoritmos avanzados de aprendizaje automático que puedan identificar características específicas de los *deepfakes* y distinguirlos de los vídeos auténticos (Simón Soler, 2023; Jacobsen & Simpson, 2023; Okolie, 2023).

El principio de estos marcos de detención es aprovechar la tecnología para localizar y prevenir la difusión de contenido manipulado. Como ejemplo encontramos el Desafío de Detección de Deepfake (DFDC) de Meta, que involucra a diversas entidades en la creación de tecnologías para identificar y prevenir el uso engañoso de *deepfakes* (Mania, 2022).

Se proponen diversas soluciones tecnológicas además del uso de *software* de detección de inteligencia artificial para identificar evidencia de manipulación en imágenes o vídeos ya mencionado. Por ejemplo, el uso de marcas de agua y tecnología *blockchain* para verificar la autenticidad del contenido digital y proporcionar un rastro de auditoría. Además, se aboga por un enfoque de "seguridad por diseño y arquitectura" para regular los *deepfakes*, donde la tecnología se diseñe considerando usos legales específicos (Okolie, 2023).

3.3.2 Regulación y propuestas legislativas

La regulación efectiva para abordar los problemas emergentes relacionados con los *deepfakes* es otro de los puntos importantes de la investigación. La carencia de leyes explícitas que aborden estos temas, la dificultad para interpretar las leyes ya existentes y la regulación de plataformas al abuso de imágenes alteradas digitalmente y *deepfakes* se repite como problemas incipientes en los estudios revisados.

Los trabajos enfatizan en la necesidad de desarrollar marcos legales y éticos sólidos para regular el uso de la inteligencia artificial. La legislación y las políticas de las plataformas existentes no pueden abordar eficazmente el daño que ha surgido y surgirá de estas técnicas (Paris & Donovan, 2019). Se subraya la importancia de la colaboración entre diferentes sectores de la sociedad, incluidos los gobiernos, la industria tecnológica, la academia y las organizaciones de derechos humanos para abordar de manera efectiva estos desafíos y mitigar los riesgos asociados.

Existe una serie de preocupaciones relacionadas con el uso de la ley como reguladora en el contexto de la proliferación de tecnología *deepfake* en la pornografía, particularmente en casos de abuso sexual basado en imágenes. Una de las principales dificultades en la aplicación efectiva de la ley radica en la identificación del perpetrador: el anonimato facilitado por la tecnología actual dificulta la atribución del *deepfake* a un individuo específico, ya que la falta de metadatos relevantes hace que sea complicado rastrear el origen del contenido. Además, las demandas civiles pueden resultar costosas y recaer sobre la víctima, resaltando la necesidad de intervención por parte de organizaciones no gubernamentales para apoyar a las víctimas (Okolie, 2023).

En ciertas jurisdicciones, las leyes que penalizan el abuso sexual basado en imágenes no abarcan los *deepfakes*. Además, en algunos lugares, la ley exige pruebas de la intención del perpetrador de dañar a la víctima, lo cual puede resultar casi imposible de demostrar en situaciones donde la víctima desconoce la creación de la imagen o cuando esta se ha realizado para uso privado o para su circulación anónima en línea. Abordar el abuso de imágenes alteradas digitalmente y *deepfakes* se vuelve aún más complicado debido a las políticas y enfoques diversos de las plataformas. Aunque algunas plataformas han tomado medidas para restringir o prohibir la compartición de contenido pornográfico de *deepfakes* y alteraciones digitales, muchas han sido lentas en responder o requieren que las víctimas se identifiquen y reporten la imagen para que se elimine (Martin, 2021; Flynn *et al.*, 2021).

En un análisis comparativo de las leyes actuales, "Legal Protection of Revenge and Deepfake Porn Victims in the European Union: Findings From a Comparative Legal Study" (Mania, 2022), se revelan diferencias significativas entre los países seleccionados para el estudio (Alemania, Bélgica, Dinamarca, España, Francia, Italia, Malta, Países Bajos y Portugal) y una falta de uniformidad en la legislación de la Unión Europea.

La mayoría de los países no aborda la pornografía de venganza como un delito sexual, sino como una infracción menor de la privacidad. En contraste, Italia se destaca por tener la legislación más sólida contra la pornografía de venganza, considerándola como un delito sexual con consecuencias severas. Por otro lado, España penaliza la distribución no autorizada de imágenes íntimas, aunque no la clasifica como delito sexual. Alemania, por su parte, opta por aplicar leyes y precedentes judiciales existentes para abordar estos casos. En cuanto a otros países como Bélgica, Dinamarca, Francia, Malta, los Países Bajos y Portugal, cuentan con disposiciones legales específicas contra la pornografía de venganza (Mania, 2022).

Considerarlo un violación menor de la privacidad y no como un delito sexual trae muchas problemáticas. No solo porque, una vez más, se centra en la víctima y no en el agresor, sino porque no reconoce la violencia estructural hacia los cuerpos feminizados. Reconocer la ciberviolencia como una forma de violencia de género es vital para hacer frente a la problemática.

Es imperativo que los países tomen medidas para establecer regulaciones adecuadas y revisen las leyes vigentes para asegurar la protección de las víctimas. Asimismo, las plataformas en línea tienen la capacidad de implementar normativas internas para abordar este problema, se insta a estas plataformas a garantizar la seguridad de sus usuarios, reforzando los controles de privacidad y tomando medidas automáticas contra el contenido abusivo que viole sus términos de servicio (Okolie, 2023).

3.3.3 Respuesta política

Por otra parte, se plantea una respuesta política para abordar la incertidumbre informativa en la era de la posverdad, promoviendo políticas y regulaciones que fomenten la veracidad y la transparencia en la información digital. Esto implica la implementación de medidas de verificación de datos y la promoción de la educación mediática para ayudar a las personas a discernir entre información falsa y verdadera (Simón Soler, 2023; Okolie, 2023).

Siguiendo con esta idea, las investigaciones consideran de suma importancia proporcionar a la población las herramientas necesarias para identificar la desinformación y cultivar una cultura de verificación de datos. Este proceso no solo implica fomentar la alfabetización mediática y digital en todas las etapas educativas, desde la infancia hasta la educación superior, sino también adaptar los contenidos según la edad y sensibilizar sobre problemáticas de género, como el uso no autorizado de imágenes para generar contenido sexual. Además, los medios de comunicación tienen un papel crucial en la difusión de campañas de concienciación y sensibilización para abordar estos temas de manera efectiva (Lucas, 2022; Simón Soler, 2023; Laffier & Rehman, 2023).

Los estudios piden una respuesta integral ante los problemas derivados por esta tecnología. Se resalta la necesidad de una colaboración estrecha entre diversos sectores de la sociedad, que abarcan desde entidades gubernamentales hasta empresas tecnológicas, instituciones académicas y organizaciones de derechos humanos, con el fin de afrontar de manera efectiva los desafíos surgidos por el avance de las tecnologías de inteligencia artificial y reducir los riesgos que conllevan (Simón Soler, 2023).

4. Discusión y conclusiones

La revisión bibliográfica ha respondido a todas las preguntas de las que partía la investigación, a partir de la búsqueda, evaluación, síntesis y análisis de artículos académicos que trataran los *deepfakes* sexuales y la violencia sexual.

Los *deepfakes* pornográficos son reconocidos como una forma de violencia sexual y presentan un sesgo de género evidente al perpetuar la sexualización y objetivación de las mujeres sin su consentimiento. Estos *deepfakes* se inscriben dentro del abuso sexual basado en imágenes (ASBI), que incluye la producción y difusión de material sexual explícito sin consentimiento. Esta forma de abuso ha migrado al espacio digital, facilitando su perpetración y causando graves daños a las víctimas.

El consentimiento juega un papel fundamental en este entorno digital, donde la falta de él es una característica central de los *deepfakes* y otros abusos sexuales en línea. La tecnología actual permite la creación de contenido sexualmente explícito a partir de imágenes de cualquier individuo sin su conocimiento o permiso, exacerbando el problema del consentimiento digital. Esta falta de consentimiento refleja una dinámica de poder desigual en línea. Los impactos de esta nueva forma de abuso sexual digital son devastadores para las víctimas, enfrentando trastornos de estrés postraumático, depresión y ansiedad como consecuencia de estas agresiones. Las motivaciones de los perpetradores van desde el placer sexual hasta la venganza y la intimidación, reflejando una combinación de deseos de control, poder y gratificación personal.

El aumento de los abusos en línea se ha visto potenciado por la Inteligencia Artificial, que ha aumentado la violación de la privacidad sexual en línea. La fácil accesibilidad a herramientas de manipulación de imágenes, como Adobe After Effects y proyectos de código abierto, posibilita la creación de *deepfakes* sexuales por parte de cualquiera con esa intención. Si bien no se le atribuye toda la responsabilidad de estos abusos a la tecnología *deepfake*, sí ha ampliado las posibilidades de esta perpetración, permitiendo la creación y difusión de contenido falso con relativa facilidad y anonimato.

Los *deepfakes* sexuales refuerzan una visión misógina y cosificadora de las mujeres, enraizada en la historia cultural de considerarlas como objetos pasivos de deseo y control, pero la tecnología en sí misma no tiene una inclinación misógina. Esta dinámica se enmarca en un contexto cultural y digital, siendo esta nueva forma de abuso sexual una continuación de problemas ya existentes en la sociedad.

En respuesta a esta problemática, se están desarrollando diversas estrategias para frenar la creación y difusión de *deepfakes* sexuales. Entre ellas se encuentran el desarrollo de métodos de detección de *deepfakes* utilizando inteligencia artificial, el establecimiento de marcos legales y éticos sólidos para regular el uso de la tecnología, y la implementación de políticas y regulaciones que promuevan la veracidad y transparencia en la información digital. También se incluye la alfabetización mediática y digital de la población para identificar la desinformación. Sin embargo, la identificación del perpetrador y la regulación de los *deepfakes* plantean desafíos significativos. Persisten desafíos en la aplicación efectiva de la ley y la colaboración entre diferentes sectores de la sociedad para abordar la cuestión de manera integral.

El uso continuo de *deepfakes* con motivaciones misóginas, especialmente para la producción y difusión de contenido sexual no consensuado, debe ser considerado como un aspecto central en cualquier análisis de esta tecnología. Los *deepfakes* plantean desafíos significativos en términos de privacidad, seguridad y equidad de género. Abordar estas preocupaciones de manera integral y colaborativa es crucial, involucrando a legisladores, empresas tecnológicas, defensores de los derechos de las mujeres y la sociedad en su conjunto. Solo a través de un enfoque conjunto se pueden encontrar soluciones efectivas y mitigar los impactos negativos de esta tecnología emergente.

5. Bibliografía

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

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

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

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

Flynn, A., Powell, A., Scott, A. J., & Cama, E. (2021). Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging form of Image-Based Sexual Abuse. *The British Journal of Criminology*, *62*(6), 1341–1358. https://doi.org/10.1093/bjc/azab111

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

Harper, C. A., Fido, D., & Petronzi, D. (2019). Delineating non-consensual sexual image offending: Towards an empirical approach. https://doi.org/10.31234/osf.io/vpydn

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

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

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

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

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

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

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

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

Roy, R., Dixit, A. K., Saxena, S., & Memoria, M. (2023). Meta-Analysis of Artificial Intelligence Solution for Prevention of Violence Against Women and Girls. 2023 International Conference on IoT, Communication and Automation Technology (ICICAT). https://doi.org/10.1109/icicat57735.2023.10263765

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

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

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

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

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

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

Cecilia Barba Arteaga (2024): Deepfakes sexuales: impacto, prevención y perspectivas de género en el entorno digital, en *Miguel Hernández Communication Journal, Vol. 15* (2), pp. 229 a 244. Universidad Miguel Hernández, UMH (Elche-Alicante). DOI: 10.21134/zt4eht31