Fronteiras: Journal of Social, Technological and Environmental Science
http://periodicos.unievangelica.edu.br/fronteiras/
ISSN 2238-8869

Article

Datificación y Violencia de Género: La Arquitectura del Silenciamiento

Maria Lorena Rossel Castagneto[1], Jorge Astudillo Muñoz[2]

[1] Doctora en Derecho. Universidad de Las Américas. ORCID: 0000-0003-4085-3000. Email: mrossel@udla.cl
[2] Doctor en Derecho. Universidad Andrés Bello. ORCID: 0000-0001-6562-1930. Email: jastudillo@unab.cl

RESUMEN
La presente investigación analiza cómo la datificación de rasgos biométricos, los algoritmos de recomendación y las estrategias de amplificación artificial han transformado la violencia de género en entornos digitales, configurando una infraestructura tecnológica que excede las categorías tradicionales de responsabilidad jurídica. La relevancia del estudio radica en que estas dinámicas intensifican las agresiones contra las mujeres y afectan la igualdad, la libertad de expresión y la participación democrática. El objetivo consiste en examinar la relación entre datificación biométrica, microsegmentación algorítmica y violencia digital, identificando sus implicancias para la regulación jurídica y la protección de los derechos fundamentales. Se emplea una metodología dogmático-jurídica, con enfoque analítico y comparado, apoyada en una revisión narrativa de literatura especializada, normativa nacional e internacional y aportes doctrinales sobre inteligencia artificial, protección de datos, plataformas digitales y violencia de género. Los resultados evidencian que la transición desde la datificación conductual hacia la biométrica posibilita nuevas formas de manipulación de la identidad, como los deepfakes, mientras que los algoritmos de recomendación, la microsegmentación y las campañas coordinadas de amplificación incrementan la difusión, personalización y permanencia del daño. Se concluye que la violencia digital contra las mujeres constituye un fenómeno estructural asociado a la arquitectura de las plataformas, lo que exige superar respuestas centradas en el agresor individual e incorporar deberes de transparencia algorítmica, diligencia reforzada y mecanismos de prevención, mitigación y reparación.
Palabras clave: Datificación, algoritmos, discriminación, ciberviolencia, género, deepfakes, autocensura, democracia.

ABSTRACT
This study examines how the datification of biometric traits, recommendation algorithms, and artificial amplification strategies have transformed gender-based violence in digital environments, giving rise to a technological infrastructure that transcends traditional categories of legal responsibility. The significance of this research lies in the fact that these dynamics intensify violence against women while undermining equality, freedom of expression, and democratic participation. The study aims to analyze the relationship between biometric datification, algorithmic microtargeting, and digital violence, identifying their implications for legal regulation and the protection of fundamental rights. To this end, it adopts a doctrinal legal methodology with an analytical and comparative approach, supported by a narrative review of specialized literature, national and international legal frameworks, and scholarly contributions on artificial intelligence, data protection, digital platforms, and gender-based violence. The findings reveal that the transition from behavioral to biometric datification has enabled new forms of identity manipulation, such as deepfakes, while recommendation algorithms, algorithmic microtargeting, and coordinated amplification campaigns increase the dissemination, personalization, and persistence of harm. The study concludes that digital violence against women is a structural phenomenon embedded in the architecture of digital platforms. Consequently, effective legal responses must move beyond approaches focused solely on the individual perpetrator and incorporate obligations of algorithmic transparency, enhanced due diligence, and mechanisms for prevention, mitigation, and redress.
Keywords: Datafication, Algorithms, Discrimination, Cyberviolence, Gender, Deepfakes, Self-censorship, Democracy

Introducción

La revolución digital contemporánea no solo ha transformado las formas y vías de comunicación, sino también algunas de las condiciones estructurales en las que se ejerce violencia en contra de las mujeres. Así, los sistemas de inteligencia artificial (IA), los algoritmos de recomendación y las plataformas digitales se han convertido en elementos centrales de la vida pública en nuestros tiempos, mediando el acceso a la información,

Submissão: 21/08/2026 Aceite: 07/10/2026 Publicação: 14/11/2026

v.15, n.3, 2026 • p. 310-323. • DOI http://dx.doi.org/10.21664/2238-8869.2026v15i3.9084

© 2021 by the authors. Esta revista oferece acesso livre imediato ao seu conteúdo, seguindo o princípio de que disponibilizar gratuitamente o conhecimento científico ao público proporciona maior democratização mundial do conhecimento. Este manuscrito é distribuído nos termos da licença Creative Commons – Atribuição - NãoComercial 4.0 Internacional (https://creativecommons.org/licenses/by-nc/4.0/legalcode), que permite reproduzir e compartilhar o material licenciado, no todo ou em parte, somente para fim não comercial; e produzir, reproduzir, e compartilhar material adaptado somente para fim não comercial.

la formación de la opinión pública y la interacción social. Al margen de los beneficios que este nuevo escenario trae consigo, esta misma infraestructura tecnológica -que no constituye un espacio neutral- puede reproducir y amplificar asimetrías de poder preexistentes, entre ellas las desigualdades estructurales que históricamente han afectado negativamente a las mujeres en el espacio público.

En este contexto, la violencia de género en entornos digitales ya no puede entenderse, analizarse y explicarse exclusivamente a partir de las categorías clásicas marcadas por una agresión (física, verbal, psicológica, económica, etc.) interpersonal, protagonizada y atribuida a un agresor identificable. De esta manera, la interacción que se produce entre la datificación de características biométricas, microsegmentación algorítmica y las estrategias de amplificación artificial -por ejemplo, campañas coordinadas de bots- pueden configurar una infraestructura digital que facilita la producción de agresiones más personalizadas, permanentes en el tiempo y escalables, dificultando la atribución jurídica de responsabilidades.

Este nuevo escenario plantea importantes desafíos regulatorios: la opacidad de los sistemas algorítmicos, la intervención simultánea de personas naturales y sistemas automatizados, además de la velocidad con que evolucionan estas formas de agresión digital, tensionan las categorías tradicionales del Derecho, que han sido pensadas y configuradas para atribuir responsabilidades a partir de actos individuales antes que de infraestructuras. En suma, limitar la respuesta estatal a la persecución penal de agresores individuales resulta insuficiente en este nuevo escenario.

En este orden de cosas, el presente estudio postula que la datificación biométrica y la arquitectura algorítmica de las plataformas digitales constituyen, en conjunto, una infraestructura que intensifica la violencia de género, confiriéndoles capacidades inéditas de reproducción, personalización y viralización del daño ocasionado a la víctima. Además, esta nueva arquitectura no solo victimiza individualmente a las mujeres expuestas a estas formas de violencia, sino que también, de alguna manera cumple una especie de función "ejemplificadora" de alcance colectivo: al naturalizar el costo de la exposición pública, puede favorecer la autocensura y el retiro progresivo de las mujeres de los espacios de deliberación democrática, socavando con ello bienes esenciales para la democracia y la dignidad humana como el pluralismo, la igualdad y la participación ciudadana.

De esta manera, el análisis de las relaciones entre tecnología, desigualdad de género y regulación jurídica, contribuye también al debate sobre la sostenibilidad social de los entornos digitales, en cuanto espacios, cuya configuración debe favorecer condiciones de igualdad, inclusión y participación democrática.

El presente estudio se estructura de la siguiente manera. Luego de dar cuenta del enfoque metodológico empleado, se procede a analizar el tránsito desde la datificación de conductas hacia la datificación de rasgos biométricos individuales, condición que posibilita el uso de herramientas como los *deepfakes* y otras formas de manipulación de la identidad digital de una persona. Posteriormente, se examina el funcionamiento de los algoritmos de recomendación y de las estrategias de amplificación artificial -como la microsegmentación y el *astroturfing*- en la configuración de la violencia de género digital. En último término, este trabajo culmina con una discusión sobre las implicancias regulatorias de este diagnóstico, particularmente en torno a la transparencia algorítmica y los deberes de diligencia de las plataformas digitales.

Metodología

La presente investigación adopta una metodología jurídico-dogmática, orientada a identificar, interpretar y sistematizar las normas, principios y categorías jurídicas relevantes para el análisis de la ciberviolencia de género en entornos digitales mediados por inteligencia artificial. Desde esta perspectiva, el estudio examina las implicancias jurídicas derivadas de la datificación biométrica, la microsegmentación algorítmica y la amplificación automatizada de contenidos, con el propósito de evaluar la suficiencia de los marcos regulatorios vigentes para la protección de los derechos fundamentales de las mujeres.

No obstante, el fenómeno analizado excede el ámbito estrictamente normativo. Como sostiene Castillo Morales (2019, 16-19), la dogmática jurídica adquiere mayor capacidad explicativa cuando incorpora herramientas provenientes de otras disciplinas que permiten comprender los procesos sociales y tecnológicos sobre los cuales opera el Derecho. En consecuencia, la investigación adopta un enfoque interdisciplinario y recurre al derecho comparado, integrando aportes del Derecho constitucional, penal y de la protección de datos, junto con contribuciones provenientes de los estudios sobre gobernanza digital, inteligencia artificial, comunicación política y plataformas digitales.

La comparación jurídica se centra en los desarrollos regulatorios de la Unión Europea y Estados Unidos, incorporando referencias puntuales a experiencias latinoamericanas, particularmente en materia de responsabilidad de plataformas digitales, transparencia algorítmica, protección de datos biométricos y violencia digital contra las mujeres. Este análisis comparado de carácter selectivo, permite identificar diferencias entre los

enfoques regulatorios examinados, evaluar sus principales fortalezas y limitaciones y formular criterios jurídicos aplicables a la protección de los derechos fundamentales frente a las nuevas formas de violencia facilitadas por la inteligencia artificial.

Las fuentes primarias analizadas comprenden instrumentos internacionales de derechos humanos, legislación nacional y comparada, documentos regulatorios sobre inteligencia artificial y protección de datos, así como una sentencia relevante de la Corte Constitucional de Colombia. Las fuentes secundarias corresponden a literatura científica especializada, informes elaborados por organismos internacionales y estudios académicos recientes sobre datificación, algoritmos, discursos de odio y violencia digital de género examinados mediante una revisión narrativa de la literatura. El tratamiento de las fuentes se realizó mediante análisis dogmático e interpretación sistemática, lo que permitió construir las categorías analíticas utilizadas para explicar la relación entre datificación biométrica, amplificación algorítmica y violencia digital de género.

Este enfoque metodológico permite responder a los objetivos de la investigación: explicar la relación existente entre datificación biométrica, arquitectura algorítmica y ciberviolencia de género, evaluar críticamente los modelos regulatorios examinados y proponer criterios jurídicos orientados a fortalecer la protección de los derechos fundamentales de las mujeres en los entornos digitales.

Resultados y Discusión

Datificación de rasgos individuales: del dato conductual al dato biométrico

La literatura sobre datificación social ha puesto históricamente el acento en el dato conductual, entendido como el registro y cuantificación de preferencias, interacciones y patrones de consumo que permiten segmentar audiencias con fines comerciales, políticos u otros. En esta línea, Kennedy et al. (2015, 1) definen la datificación como el proceso mediante el cual aspectos del mundo social previamente no cuantificados son convertidos en datos, comprendiendo tanto información demográfica o de perfilamiento como metadatos conductuales -geolocalización, patrones de interacción o marcas temporales de uso, entre otros-. Esta concepción ha servido para explicar cómo la recopilación masiva de datos permite conocer, predecir e influir sobre el comportamiento humano (Arenas 2024, 121-124; Ábrego Molina y Flores Mérida 2021, 214-216).

La progresiva incorporación de datos biométricos a estos procesos de datificación plantea, sin embargo, una dimensión adicional: la utilización de características físicas y fisiológicas como elementos de identificación y representación de personas. Esta atribución de valor identificatorio privilegiado al dato biométrico no es casual, sino que responde a lo que Van Dijck (2014, 201) denomina "dataísmo": la creencia, en la capacidad de los datos para representar objetivamente la realidad social, que permite interpretar datos agregados -o, en este caso, rasgos biofísicos individualizados- como acceso directo y objetivo a la verdad sobre un sujeto. Trasladada al plano biométrico, esta lógica "dataísta" contribuye a explicar por qué el rostro o la voz sintetizados mediante IA generativa resultan socialmente verosímiles como representación auténtica de la víctima: el dataísmo naturaliza la idea de que el dato biofísico, una vez extraído y recompuesto algorítmicamente, sigue siendo un correlato fiel de la identidad real de las personas, a pesar de que el contenido producido sea íntegramente falso (Chesney y Citron 2019, 1757).

Desde una perspectiva más amplia, Gendler (2024, 122-124) sitúa estos procesos dentro de una genealogía de poder biopolítico, apoyada en los desarrollos de la medicina, la biología y la economía política, en la que los datos han sido utilizados para modelar "subjetividades, cuerpos, edificaciones y gestiones de acciones y comportamientos". Sin embargo, tanto la definición clásica de Kennedy *et al.* como la caracterización histórica propuesta por Gendler fueron concebidas principalmente para explicar la datificación del comportamiento individual y colectivo aunque no agotan las implicancias específicas de la utilización contemporánea de rasgos biométricos. Ese marco resulta hoy insuficiente para comprender una transformación cualitativamente distinta: aquella que ya no recae sobre lo que el sujeto hace, sino también sobre lo que el sujeto es. En efecto, la progresiva captura y tratamiento del rostro, la voz y otros rasgos biométricos desplaza el objeto de la datificación desde la conducta hacia la propia identidad biofísica de las personas.

En el RGPD, la protección especial comprende los datos biométricos tratados para identificar de manera unívoca a una persona física. Desde una perspectiva jurídica, esta distinción no es meramente descriptiva, sino que posee consecuencias normativas de primer orden. Los marcos de protección de datos personales -desde el Reglamento General de Protección de Datos europeo hasta la reciente Ley 21.719 en Chile- reconocen a los datos biométricos como una categoría sensible o especial, sometida a un régimen de tutela reforzada precisamente porque su tratamiento no autorizado compromete no solo la privacidad informacional, sino también derechos vinculados a la identidad del titular (Reglamento [UE] 2016/679, art. 9; Ley N.º 21.719, 2024). Por ello, Álvarez Buján (2023, 185-188) advierte que la utilización de sistemas de inteligencia artificial exige

reforzar las garantías de tutela judicial efectiva y autodeterminación informativa, especialmente cuando las decisiones automatizadas inciden sobre derechos fundamentales (Rebollo Delgado 2023, 122).

La relevancia de esta transformación excede, sin embargo, el ámbito de la protección de datos personales. La datificación permanente de rasgos biométricos modifica las condiciones tradicionales del anonimato en el espacio público. Como sostiene Ramos-Zaga (2026, 14), las tecnologías de reconocimiento facial permiten vincular la presencia física con identidades digitales y bases de datos personales, debilitando la expectativa de anonimato asociada a la presencia transitoria en espacios públicos. En este contexto, la biometrización puede convertir la identidad física en un objeto de registro continuo, clasificación y reutilización tecnológica.

Esta segunda forma de datificación —la del rasgo individual antes que la del comportamiento agregado— constituye uno de los presupuestos tecnológicos que puede facilitar la elaboración de los llamados *deepfakes*: contenidos audiovisuales hiperrealistas generados mediante el entrenamiento de modelos de aprendizaje automático sobre bancos de imágenes, videos o grabaciones de voz de una persona determinada. Mientras la microsegmentación pretende identificar a quién persuadir, la datificación biométrica permite decidir sobre quién construir el mensaje. El objeto de la intervención algorítmica deja de ser únicamente el receptor y pasa a ser también la identidad digital de la víctima. A diferencia de la microsegmentación conductual, que opera sobre la audiencia que recibe un mensaje, la datificación biométrica opera sobre el sujeto representado, transformando su imagen y su voz en un recurso manipulable y reproducible por terceros.

Esta evolución puede leerse, además, como una manifestación del denominado capitalismo de vigilancia. Del mismo modo en que el dato conductual es transformado en un "activo de vigilancia" y en materia prima para la producción de "futuros conductuales" comercializables (Lucas 2020, 51), el rostro, la voz y los demás rasgos biométricos adquieren un nuevo valor económico como recursos susceptibles de extracción, almacenamiento y reutilización. La datificación biométrica puede extender así la lógica extractiva propia del capitalismo de vigilancia desde el comportamiento hacia la identidad física de las personas. En el contexto de la violencia digital de género, sin embargo, esa instrumentalización ya no persigue únicamente finalidades comerciales, sino también la producción deliberada de daño mediante la manipulación de la identidad digital de las víctimas, condicionando el comportamiento de los individuos y restringiendo el ejercicio efectivo de derechos fundamentales (G'sell 2025, 8-12).

El tránsito desde la datificación conductual hacia la datificación biométrica supone también una transformación del objeto mismo del poder digital. Mientras la primera busca predecir y modificar conductas futuras mediante inferencias probabilísticas sobre preferencias individuales, la segunda permite intervenir directamente sobre los atributos constitutivos de la identidad personal. El rostro, la voz o la gestualidad dejan de ser únicamente elementos identificatorios para convertirse en insumos susceptibles de extracción, reproducción y manipulación mediante inteligencia artificial. En consecuencia, los rasgos físicos utilizados en procesos de identificación o representación digital pueden servir no solo para identificar a una persona, sino también para generar representaciones sintéticas de ella.

Esta transformación tiene una consecuencia normativa que Ramos-Zaga (2026, 20) formula como criterio de política pública: la necesidad de consagrar "la inalienabilidad de la huella biométrica, elevando el anonimato en espacios públicos a la categoría de derecho fundamental no negociable". Esta propuesta de inalienabilidad resulta plenamente coherente con la distinción jurídica entre dato conductual y dato biométrico: si el primero admite regímenes de consentimiento y portabilidad -propios de la lógica contractual del dato personal-, el segundo, por constituir la identidad biofísica misma del sujeto, exige un estatuto de protección reforzada que, conforme a dicha propuesta, no puede quedar sujeto a la disponibilidad contractual del titular, precisamente porque su apropiación no consentida (como puede ocurrir con la elaboración de los *deepfakes*) no vulnera solamente la privacidad informacional, sino la integridad de la identidad misma.

La biometrización no crea por sí sola nuevas formas de misoginia ni de violencia contra las mujeres. Proporciona, sin embargo, una infraestructura tecnológica capaz de intensificar violencias históricamente existentes mediante mecanismos inéditos de reproducción, automatización y viralización del daño. En los entornos digitales, las plataformas no solo pueden amplificar desigualdades estructurales preexistentes, sino que facilitan nuevas modalidades de difusión y persistencia de las agresiones, lo que puede incrementar su alcance y permanencia. En este sentido, la datificación biométrica constituye un presupuesto técnico que potencia formas tradicionales de violencia de género, dotándolas de capacidades de manipulación, escalabilidad y circulación que adquieren nuevas dimensiones en los entornos digitales (Avendaño Paredes *et al.* 2025, 6-8; Viña Castillo *et al.* 2025, 2069-2070).

Esta mutación tecnológica tiene especial relevancia para la violencia digital basada en género. Si la microsegmentación clásica persigue identificar a las audiencias más susceptibles de recibir determinados mensajes, la datificación biométrica desplaza el foco hacia la propia víctima, cuyo cuerpo digitalizado pasa a

constituir el objeto de la manipulación. La extracción y reutilización de imágenes, voces y otros rasgos personales permite producir representaciones falsas altamente verosímiles -como ocurre con los deepfakes sexuales- que posteriormente son insertadas en ecosistemas algorítmicos capaces de amplificar discursos de odio, campañas de hostigamiento y violencia coordinada.

En definitiva, la datificación biométrica transforma la identidad física de las personas en un objeto susceptible de extracción, clasificación y manipulación algorítmica. Este desplazamiento desde el dato conductual hacia el dato biométrico no solo redefine el alcance de la protección jurídica de los datos personales, sino que contribuye a configurar la infraestructura tecnológica sobre la cual operan posteriormente la microsegmentación, los sistemas de recomendación y las estrategias de amplificación que caracterizan la violencia digital contemporánea. Comprender esta transformación resulta indispensable para analizar cómo la arquitectura algorítmica convierte esa materia prima en mecanismos de producción, distribución y escalamiento del daño contra las mujeres, cuestión que se examina en el apartado siguiente.

Algoritmos, Plataformas y Discursos de Odio: La Ciberviolencia de Género en el Ecosistema Digital

El análisis desarrollado en el apartado precedente permitió explicar cómo la incorporación de la datificación biométrica a los procesos de datificación conductual ha transformado la identidad de las personas en un objeto susceptible de extracción, clasificación y manipulación algorítmica. No obstante, la explicación de la ciberviolencia de género no se agota en la obtención y reutilización de datos biométricos. Resulta igualmente necesario analizar la infraestructura tecnológica que determina la circulación, visibilidad y amplificación de esos datos y de los contenidos construidos a partir de ellos. Ello exige desplazar el análisis desde la lógica de la datificación hacia la lógica de la distribución algorítmica de la atención (Rubio *et al.* 2024, 31-34).

Las plataformas digitales y sus sistemas de recomendación constituyen el núcleo de esa segunda dimensión del problema. Si la datificación biométrica proporciona una de las fuentes de información sobre la que pueden construirse contenidos sintéticos, es la arquitectura algorítmica la que influye en qué contenidos adquieren visibilidad, a qué audiencias son dirigidos y con qué intensidad son amplificados (Mota 2023, 14-18). Es precisamente en la interacción entre datificación biométrica, microsegmentación y sistemas de recomendación donde se configura la arquitectura contemporánea de la ciberviolencia de género y donde los discursos de odio contra las mujeres dejan de depender exclusivamente de la acción de un agresor individual para convertirse en fenómenos de difusión y amplificación estructural.

La revolución digital prometió un espacio de comunicación horizontal y democratización del conocimiento; sin embargo, el ciberespacio ha demostrado estar lejos de constituir un entorno neutral. Como advierten Avendaño Paredes *et al.* (2025, 14), el ecosistema digital reproduce y amplifica las asimetrías de poder y las violencias estructurales presentes en el mundo offline, consolidando dinámicas patriarcales que hostigan, silencian y subordinan a las mujeres. En este contexto, la intersección entre la perfilación algorítmica, la microsegmentación y la propagación automatizada de discursos de odio ha configurado un escenario de especial vulnerabilidad que plantea desafíos inéditos para el Derecho y para la política criminal.

Las plataformas digitales han dejado de desempeñar un papel meramente instrumental para convertirse en infraestructuras que organizan la circulación de la información y condicionan la formación de la opinión pública. En el marco del denominado capitalismo de vigilancia, la extracción sistemática de datos conductuales y biométricos constituye el fundamento económico de modelos de negocio basados en la predicción y modulación del comportamiento humano (Ramos-Zaga 2026, 11-13). En consecuencia, los algoritmos de recomendación no solo determinan qué contenido visualiza cada usuario, sino que también pueden condicionar la intensidad de su exposición, moldear preferencias y favorecer procesos de polarización (Araujo Pulido *et al.* 2026, 11; Ramos-Zaga 2026, 5; Viña Castillo *et al.* 2025, 2069-2070).

Esta lógica de priorización no constituye una consecuencia accidental del diseño tecnológico, sino una manifestación del modelo económico que estructura el funcionamiento de las plataformas digitales. En el marco de la teoría del capitalismo de vigilancia desarrollada por Zuboff, el denominado "instrumentarismo" supone la utilización sistemática de los datos y de la conducta humana con fines de predicción, modificación, monetización y control (Lucas 2020, 51). En este contexto, la atención de las personas deja de ser un mero recurso cognitivo para convertirse en un activo comercializable, de modo que los algoritmos tienden a privilegiar aquellos contenidos capaces de generar mayores niveles de interacción y permanencia en las plataformas. Los contenidos emocionalmente intensos -entre ellos los discursos de odio, la misoginia y otras formas de violencia dirigidas contra las mujeres- pueden resultar especialmente funcionales a esa lógica, en la medida en que incrementan la producción de datos conductuales y fortalecen la capacidad predictiva de los sistemas algorítmicos. En consecuencia, la amplificación de este tipo de contenidos no constituye un efecto colateral susceptible de corregirse mediante ajustes técnicos aislados, sino que debe examinarse también en

relación con el modelo de negocio que sostiene la economía de la vigilancia y la arquitectura contemporánea de recomendación.

En este escenario, la microsegmentación ha dejado de ser una herramienta propia del marketing destinada a personalizar mensajes comerciales. Actualmente constituye uno de los principales mecanismos mediante los cuales las plataformas determinan qué contenidos serán mostrados, a qué audiencias y con qué grado de intensidad serán difundidos. De este modo, la microsegmentación deja de ser únicamente una técnica de publicidad personalizada para convertirse en una infraestructura de distribución diferencial del daño digital. Los algoritmos de recomendación de plataformas como Instagram, X o TikTok pueden favorecer la viralización de contenidos altamente emocionales, polémicos o vejatorios, generando ciclos de retroalimentación que intensifican la sexualización, la misoginia y otras formas de violencia digital (Celuch *et al.* 2023, 625; Gobierno de España, 2026, 6). A ello se suma que los sistemas de inteligencia artificial, al entrenarse con datos que reflejan desigualdades históricas, pueden reproducir sesgos estructurales que perpetúan estereotipos discriminatorios y profundizan fenómenos de discriminación algorítmica (Ferrante 2021, 27-36; Centro de Derechos Humanos UDP 2024, 368).

En consecuencia, la violencia de género en el entorno digital ya no puede comprenderse exclusivamente como el resultado de la conducta de un agresor individual. Se inserta, más bien, en una auténtica arquitectura del daño digital, en la que convergen agresores humanos, plataformas tecnológicas, sistemas algorítmicos y modelos económicos sustentados en la economía de la atención y el capitalismo de vigilancia (Ramos-Zaga 2026, 24; Viña Castillo *et al.* 2025, 2070). La datificación constituye uno de los presupuestos técnicos de esa arquitectura, pues convierte la actividad humana —y, como se explicó en el apartado precedente, también los rasgos biométricos que individualizan a las personas— en datos susceptibles de clasificación, predicción y reutilización algorítmica. En este contexto, los algoritmos dejan de ser instrumentos neutrales para operar como verdaderos dispositivos sociotécnicos que distribuyen la atención pública, jerarquizan la visibilidad de determinados contenidos y pueden amplificar dinámicas de confrontación y hostigamiento (Araujo Pulido *et al.* 2026, 11).

A través de la microsegmentación y del perfilamiento algorítmico, los sistemas de procesamiento masivo utilizan *proxies* —variables aparentemente neutrales, como códigos postales, patrones de consumo o hábitos de navegación— que, en la práctica, mantienen una estrecha correlación con categorías protegidas como el género, la raza o la clase social (Ramos-Zaga 2026, 24). Ello puede producir un impacto dispar (*disparate impact*), reproduciendo y amplificando desigualdades estructurales preexistentes mediante decisiones automatizadas que afectan de manera desproporcionada a determinados grupos (Ramos-Zaga 2026, 24 y 25; Sosa Huapaya 2024, 288-290). A su vez, al estar diseñados para maximizar el *engagement*, los algoritmos privilegian contenidos polémicos, violentos o sexualizados, incrementando la exposición de las mujeres a dinámicas de hostigamiento y dando lugar a lo que parte de la doctrina identifica como violencia algorítmica (Araujo Pulido *et al.* 2026, 11; Viña Castillo *et al.* 2025, 2069-2070). Sobre esta infraestructura tecnológica se articulan, además, estrategias coordinadas de manipulación como el astroturfing, mediante las cuales bots, cuentas falsas y redes de *trolls* fabrican artificialmente la apariencia de apoyo o rechazo social para amplificar campañas de hostigamiento y discursos de odio. La coordinación encubierta constituye precisamente el elemento distintivo de estas prácticas, pues busca presentar como expresión espontánea de la ciudadanía lo que en realidad corresponde a acciones deliberadamente organizadas (Tasente y Popescu 2025, 6-8; García-Estévez *et al.* 2025, 173-175).

La infraestructura de datos y los sistemas algorítmicos no poseen, por sí mismos, un significado intrínsecamente antidemocrático. Como muestran Milan y Beraldo (2024, 272), estas mismas herramientas pueden ser reapropiadas por la sociedad civil para favorecer procesos de participación democrática y acción colectiva. En consecuencia, la diferencia entre una infraestructura de datos orientada al fortalecimiento de la deliberación pública y otra destinada a amplificar discursos de odio no reside exclusivamente en la tecnología empleada, sino también en las finalidades que orientan su utilización y en las dinámicas de poder que reproduce. Bajo esta perspectiva, la arquitectura algorítmica constituye una infraestructura ambivalente, capaz tanto de favorecer la agencia colectiva como de intensificar formas de manipulación y violencia digital. Precisamente esta segunda dimensión será analizada en el apartado siguiente, mediante el examen de las principales manifestaciones de ciberviolencia de género potenciadas por la microsegmentación algorítmica.

Manifestaciones de la Violencia Potenciadas por la Microsegmentación

La arquitectura algorítmica descrita anteriormente no solo condiciona la circulación de contenidos, sino que también puede transformar las modalidades mediante las cuales se ejerce la violencia de género en entornos digitales. La combinación entre microsegmentación, inteligencia artificial y sistemas automatizados de recomendación permite que las agresiones sean cada vez más personalizadas, persistentes y escalables,

afectando especialmente a mujeres con alta exposición pública, como periodistas, activistas o representantes políticas (Araujo Pulido *et al.* 2026, 4-11; Viña Castillo *et al.* 2025, 2071).

Estas transformaciones se manifiestan, principalmente, en tres mecanismos interrelacionados mediante los cuales la arquitectura algorítmica intensifica la violencia de género en entornos digitales: (i) la automatización de las agresiones; (ii) la personalización del daño mediante la microsegmentación; y (iii) la manipulación del ecosistema informativo a través de mecanismos de amplificación artificial y falso consenso.

En primer lugar, la automatización de las agresiones constituye una de las principales manifestaciones de la violencia digital contemporánea. La inteligencia artificial generativa ha facilitado la creación de *deepfakes* de contenido sexual no consentido, cuya difusión puede ser replicada masivamente mediante sistemas automatizados, ampliando considerablemente el alcance del daño y dificultando la identificación de sus responsables (Viña Castillo et al. 2025, 2070). A ello se suman campañas coordinadas de hostigamiento ejecutadas mediante granjas de *trolls* (*troll centers*) y cuentas automatizadas (*bots*), que generan oleadas de insultos misóginos, discursos transfóbicos y amenazas dirigidas contra mujeres con alta exposición pública, como periodistas, activistas y representantes políticas, con el propósito de desacreditar su actuación y forzar su retiro del debate público (Araujo Pulido *et al.* 2026, 4; Avendaño Paredes *et al.* 2025, 32). La intervención combinada de actores humanos y sistemas automatizados puede dificultar la identificación de las conductas individuales y la atribución de responsabilidades, configurando una violencia descentralizada, persistente y particularmente difícil de investigar desde la perspectiva jurídica (Viña Castillo *et al.* 2025, 2075).

En segundo lugar, la personalización del daño permite dirigir ataques específicos contra mujeres previamente perfiladas mediante técnicas de microsegmentación y tratamiento masivo de datos. A partir de información obtenida de la actividad digital de las víctimas, las plataformas y los agresores pueden identificar perfiles particularmente vulnerables y diseñar campañas dirigidas de hostigamiento. En este contexto, prácticas como el *doxxing* —la divulgación no autorizada de datos personales—, la difusión de información falsa o descontextualizada y las campañas de desprestigio se orientan a afectar la reputación profesional, la credibilidad pública y la seguridad personal de las víctimas (García 2026, 4; Avendaño Paredes *et al.* 2025, 27). Estas agresiones suelen centrarse en estereotipos de género, cuestionando la apariencia física, la sexualidad o la capacidad profesional de las mujeres como mecanismos de silenciamiento y exclusión del espacio público (Avendaño Paredes *et al.* 2025, 20).

En tercer lugar, la manipulación del ecosistema informativo amplifica y legitima la violencia mediante estrategias coordinadas que simulan consensos sociales inexistentes. Entre ellas destaca el astroturfing, técnica de manipulación que utiliza cuentas falsas, *sock puppets* y redes de bots para aparentar un apoyo ciudadano espontáneo cuando, en realidad, responde a campañas organizadas de desinformación (García-Estévez *et al.* 2025, 161). A través de estos mecanismos se difunden de forma coordinada discursos de odio y narrativas antifeministas, creando la falsa percepción de que existe un rechazo amplio y socialmente compartido hacia las reivindicaciones de igualdad o hacia determinadas mujeres por razón de su género (García-Estévez *et al.* 2025, 162). Esta apariencia de consenso no solo amplifica el alcance de los mensajes violentos, sino que también puede contribuir a normalizar la misoginia, reforzar los procesos de polarización y erosionar las condiciones de una deliberación democrática basada en información auténtica.

Estos tres mecanismos —la automatización de las agresiones, la personalización del daño y la manipulación del ecosistema informativo— no operan de manera aislada, sino que se retroalimentan mutuamente. La datificación biométrica puede proporcionar la información utilizada para construir contenidos sintéticos; la microsegmentación permite identificar las audiencias y víctimas potenciales; los sistemas automatizados amplifican su circulación; y las campañas coordinadas de manipulación refuerzan artificialmente su legitimidad y alcance. El resultado es una infraestructura digital capaz de producir daños persistentes, escalables y de difícil atribución jurídica, que plantea desafíos adicionales a los modelos tradicionales de responsabilidad por violencia interpersonal.

De este modo, las manifestaciones descritas no producen únicamente afectaciones individuales sobre las víctimas, sino que proyectan sus efectos sobre el funcionamiento mismo del espacio público digital. Cuando los *deepfakes,* el acoso coordinado, el *doxxing* o las campañas de *astroturfing* buscan desacreditar, intimidar o expulsar a las mujeres del debate público, no solo lesionan derechos como la dignidad, la igualdad, la integridad personal o la autodeterminación informativa, sino que también alteran las condiciones bajo las cuales se forma la opinión pública en los entornos digitales (Reis y Campos 2026, 189). En este sentido, la Carta de Derechos Digitales española reconoce expresamente derechos de naturaleza política, entre ellos el derecho a recibir libremente información y el derecho a la participación política por medios digitales (Domínguez Álvarez 2022, 604-605). La manipulación artificial del consenso mediante estrategias de astroturfing dirigidas contra mujeres compromete directamente estos bienes jurídicos, al distorsionar los procesos de deliberación pública, debilitar

el pluralismo democrático y favorecer el silenciamiento de voces auténticas. Desde esta perspectiva, la ciberviolencia de género trasciende la esfera de la afectación individual para convertirse también en un problema estructural de calidad democrática, en la medida en que condiciona quiénes pueden participar en el debate público, en qué condiciones lo hacen y qué discursos alcanzan visibilidad dentro del ecosistema digital (Ron Zambrano y Rodríguez Jaramillo 2025, 49-75).

Las consecuencias psicosociales de esta arquitectura de la violencia son profundas y persistentes. Las mujeres afectadas pueden experimentar ansiedad, aislamiento, afectaciones a su identidad y daños a su reputación personal y profesional, lo que puede favorecer procesos de autocensura y, en ciertos casos, su progresiva exclusión de los espacios de deliberación pública y participación democrática (Araujo Pulido *et al.* 2026, 11; Viña Castillo *et al.* 2025, 2077; Lafont 2025, 22-26; Lazard *et al.* 2025, 10).

Desde la perspectiva jurídica, esta arquitectura tecnológica evidencia importantes desafíos regulatorios. La opacidad de los algoritmos —las denominadas cajas negras—, la intervención simultánea de actores humanos y sistemas automatizados, así como la rapidez con que evolucionan estas formas de agresión, dificultan la atribución de responsabilidad y la adopción de respuestas eficaces (Araujo Pulido *et al.* 2026, 11). En este contexto, limitar la intervención estatal a la persecución penal de los agresores individuales resulta insuficiente. Una respuesta regulatoria adecuada requiere incorporar obligaciones de diligencia para las plataformas digitales, fortalecer los estándares de transparencia algorítmica y establecer mecanismos eficaces de prevención, detección y mitigación de aquellos sistemas de recomendación que contribuyen a amplificar y perpetuar la violencia de género en línea (Torres-Zambrano y Morillo-Puente 2025, 58; Viña Castillo *et al.* 2025, 2070).

El Silenciamiento y la Autocensura como Efecto Disciplinador

Los mecanismos analizados en el apartado precedente —la automatización de las agresiones, la personalización del daño y la manipulación del ecosistema informativo— no constituyen un fin en sí mismos, sino que pueden converger en un resultado común: alterar las condiciones de participación de las mujeres en el espacio público digital (Barlow y Awan 2016, 2 y 3; Noelle-Neumann1974, 43-51; Neuwirth *et al.* 2007, 450-468). Desde esta perspectiva, el silenciamiento y la autocensura no pueden entenderse únicamente como efectos colaterales de la arquitectura algorítmica, sino también como una consecuencia estructural del modelo de funcionamiento de las plataformas digitales. En los términos de Zuboff (2020, 438), esta dinámica puede interpretarse a partir del instrumentarismo, entendido como la utilización sistemática de los datos y de la conducta humana con fines de predicción, modificación, monetización y control. Cuando las agresiones digitales inducen a las mujeres a retraerse del debate público, limitar su participación o abandonar determinados espacios digitales, la violencia deja de operar únicamente como una agresión individual para convertirse en un mecanismo de disciplinamiento del espacio público (Araujo Pulido *et al.* 2026, 11). En este contexto, el miedo deja de ser únicamente una reacción subjetiva frente a la violencia para transformarse en un mecanismo estructural de regulación de la expresión y de la participación democrática (Araujo Pulido *et al.* 2026, 11).

De este modo, la arquitectura algorítmica no solo puede amplificar contenidos lesivos, sino que también puede modificar las condiciones bajo las cuales las mujeres ejercen sus derechos de expresión, participación e incidencia en los entornos digitales. El silenciamiento y la autocensura dejan así de ser respuestas exclusivamente individuales frente a la violencia para convertirse en indicadores del impacto estructural que las tecnologías de recomendación y amplificación pueden ejercer sobre el ejercicio efectivo de los derechos fundamentales en el espacio público digital.

La primera manifestación de este efecto disciplinador es la autocensura. Frente al riesgo de sufrir campañas de hostigamiento, humillación pública o amenazas de violencia física y sexual, muchas mujeres modifican voluntariamente su comportamiento digital. Ello se traduce en la reducción de sus publicaciones, la limitación de sus interacciones, el abandono de determinados temas considerados conflictivos o el cierre parcial de sus perfiles en redes sociales (Avendaño Paredes *et al.* 2025, 37). La autocensura deja así de ser una decisión plenamente libre para convertirse en una respuesta adaptativa frente a un entorno comunicativo hostil.

Esta caracterización requiere, sin embargo, una precisión conceptual. La autocensura no puede interpretarse como el resultado de una decisión plenamente libre, ni tampoco como la simple negación de la capacidad de agencia de las víctimas. Como señalan Kennedy *et al.* (2015, 2), la relación entre estructura y agencia posee un carácter dialéctico: las estructuras condicionan las posibilidades de acción de los sujetos, pero estos continúan actuando dentro de ellas e, incluso, resistiéndolas.

Aplicada a la ciberviolencia de género, esta perspectiva permite comprender que la autocensura constituye simultáneamente una respuesta individual frente a la violencia y el resultado de un entorno algorítmico que restringe progresivamente las condiciones para una participación segura e igualitaria en el espacio público digital. En consecuencia, el silencio no representa una elección plenamente autónoma, sino una forma de agencia

ejercida bajo condiciones estructurales que reducen significativamente el margen de opciones disponibles para las mujeres. Desde esta perspectiva, la autocensura deja de ser una reacción meramente individual para convertirse en un indicador del impacto que la arquitectura algorítmica ejerce sobre el ejercicio efectivo de la libertad de expresión, la participación democrática y el derecho de las mujeres a intervenir en igualdad de condiciones en los entornos digitales (McGlynn *et al.* 2017, 26-28; Jane 2017, 13).

La autocensura puede derivar en un progresivo retiro de la esfera pública digital. La exposición continuada a campañas de violencia y desprestigio puede provocar que algunas mujeres abandonen plataformas digitales o reduzcan significativamente su participación en ellas, con consecuencias directas sobre sus oportunidades profesionales, académicas y políticas (Avendaño Paredes *et al.* 2025, 38). Este fenómeno adquiere especial gravedad en el ámbito político, donde el acoso digital dirigido contra candidatas, periodistas o defensoras de derechos humanos restringe la pluralidad del debate democrático y contribuye a reproducir relaciones estructurales de desigualdad de género (Araujo Pulido *et al.* 2026, 11).

El retiro de las mujeres del espacio público digital admite, además, una lectura desde la teoría de los movimientos sociales datificados. Milan y Beraldo (2024, 277) sostienen que determinados actores sociales pueden optar por retirarse parcialmente de las dinámicas de datificación como una estrategia de resistencia frente a los mecanismos de vigilancia. No obstante, los propios autores advierten que dicha retirada no implica escapar de la lógica de la datificación, sino que pone de manifiesto el alcance de las estructuras que la sostienen. Aplicada a la ciberviolencia de género, esta reflexión permite comprender que el abandono de determinados espacios digitales por parte de las mujeres no constituye una verdadera salida de la arquitectura algorítmica, sino una manifestación de su capacidad para condicionar las formas de participación en el espacio público.

Este efecto disciplinador se ve reforzado por la normalización social de la violencia digital. Con frecuencia, las agresiones son minimizadas bajo la idea de que "solo ocurren en internet" o justificadas mediante una comprensión absoluta de la libertad de expresión (Torres-Zambrano y Morillo-Puente 2025, 58). Sin embargo, la reiteración y viralización de los discursos de odio puede favorecer un proceso de legitimación social que convierte la violencia en un costo aparentemente inevitable para las mujeres que participan activamente en los espacios digitales (Avendaño Paredes *et al.* 2025, 14).

En consecuencia, el troleo de género y los discursos de odio no constituyen meros excesos comunicativos propios de las redes sociales, sino que pueden operar como mecanismos de control social que buscan disciplinar la participación de las mujeres mediante la producción de miedo, autocensura y exclusión del debate público. Desde esta perspectiva, la violencia digital no solo vulnera derechos individuales, sino que compromete bienes colectivos esenciales para una sociedad democrática, como el pluralismo, la igualdad en la participación política y la libertad de expresión ejercida en condiciones de seguridad (García 2026, 5; Araujo Pulido *et al.* 2026, 11).

El análisis desarrollado permite reconstruir una arquitectura integrada de producción y amplificación de la ciberviolencia de género. En ella, la datificación biométrica puede proporcionar información susceptible de manipulación (1); la infraestructura algorítmica de las plataformas determina su circulación, distribución diferencial y amplificación (2); estos procesos se materializan en mecanismos concretos de violencia automatizada, personalizada y coordinada (3); y, finalmente, pueden producir un efecto estructural de silenciamiento y autocensura que altera las condiciones de participación de las mujeres en el espacio público digital (4).

Esta arquitectura no solo compromete derechos fundamentales de las víctimas individualmente consideradas, sino que también incide sobre las condiciones estructurales de la deliberación democrática. Al restringir el ejercicio efectivo de la libertad de expresión, la igualdad y la participación política de las mujeres en los entornos digitales, la ciberviolencia de género afecta el pluralismo y la calidad del debate público. En consecuencia, su tratamiento jurídico no puede agotarse en los mecanismos tradicionales de protección de datos personales o de responsabilidad individual, sino que exige una respuesta regulatoria que considere conjuntamente la datificación biométrica, la arquitectura algorítmica y el papel que desempeñan las plataformas digitales en la producción, amplificación y persistencia del daño. La respuesta del Derecho debe orientarse, por tanto, no solo a reparar el daño una vez producido, sino también a intervenir sobre las condiciones estructurales que permiten su producción y reproducción en los ecosistemas digitales.

Desafíos regulatorios frente a la arquitectura algorítmica de la ciberviolencia de género

El análisis precedente demuestra que la ciberviolencia de género potenciada por la inteligencia artificial constituye un fenómeno estructural cuya explicación excede las categorías tradicionales del derecho penal o de la protección de datos personales. La interacción entre datificación biométrica, arquitectura algorítmica y sistemas de recomendación revela que el daño no se produce únicamente por la actuación individual del agresor, sino también por el funcionamiento de plataformas que pueden contribuir a amplificar, distribuir y prolongar

la circulación de contenidos lesivos. En consecuencia, una respuesta regulatoria eficaz debe intervenir sobre cada uno de estos niveles, combinando mecanismos de responsabilidad, transparencia y sanción, sin desconocer las exigencias derivadas de la libertad de expresión y demás derechos fundamentales.

Responsabilidad de plataformas e intermediarios

Un primer eje regulatorio consiste en redefinir el régimen jurídico de responsabilidad de las plataformas digitales. La atribución exclusiva de responsabilidad al autor material del contenido resulta insuficiente frente a fenómenos caracterizados por el anonimato, la automatización y la rápida propagación de contenidos mediante sistemas algorítmicos. Como advierte Balkin (2012, 50), en muchos casos resulta considerablemente más eficaz dirigir las acciones contra los intermediarios que permiten la difusión de contenidos ilícitos que perseguir individualmente a quienes los generan.

En el derecho comparado pueden identificarse dos modelos regulatorios. El primero corresponde al europeo, consolidado mediante la Ley de Servicios Digitales (*Digital Services Act,* Unión Europea 2022), que impone deberes diferenciados de diligencia a las plataformas, mecanismos de notificación y retirada de contenidos, procedimientos de reclamación para los usuarios y obligaciones de auditoría respecto de sus sistemas algorítmicos. El segundo corresponde al modelo norteamericano, construido sobre la Sección 230 de la *Communications Decency Act* de 1996, que privilegia la innovación y la autorregulación al excluir, como regla general, la responsabilidad de las plataformas por contenidos generados por terceros.

Desde la perspectiva desarrollada en este trabajo, el modelo europeo ofrece mayores herramientas para enfrentar la ciberviolencia de género, precisamente porque establece obligaciones para las plataformas que van más allá de la responsabilidad por contenidos de terceros y alcanzan determinados riesgos asociados al funcionamiento de sus servicios y sistemas de recomendación.

Transparencia algorítmica

Un segundo eje regulatorio consiste en fortalecer los mecanismos de transparencia algorítmica. Si la arquitectura del daño digital depende del funcionamiento de algoritmos que seleccionan, jerarquizan y recomiendan contenidos, su funcionamiento no puede permanecer completamente opaco al control público.

Como sostiene Pauner Chulvi (2023, 123), la transparencia algorítmica cumple una doble función. Desde una perspectiva externa, permite controlar la existencia de errores, discriminaciones y sesgos algorítmicos, así como garantizar el ejercicio de los derechos de acceso e información. Desde una dimensión interna, facilita la revisión técnica de los sistemas automatizados por parte de programadores, analistas y auditores independientes.

Esta comprensión ha sido recogida también por la Corte Constitucional de Colombia, al sostener que la transparencia algorítmica constituye una garantía indispensable en una sociedad democrática, pues solo mediante la disponibilidad de información resulta posible controlar el impacto de los sistemas automatizados sobre la autonomía, la igualdad y la dignidad de las personas (Corte Constitucional de Colombia 2025, Finol Romero y Ávila Hernández 2026, 310).

Un modelo regulatorio integral

Finalmente, la respuesta jurídica no puede descansar exclusivamente en la expansión del derecho penal. La sanción penal continúa siendo necesaria para las manifestaciones más graves de violencia digital, pero debe operar conforme a los principios de legalidad, proporcionalidad y ultima ratio. Paralelamente, resulta indispensable fortalecer las respuestas provenientes del derecho administrativo y civil, mediante deberes reforzados de diligencia para las plataformas, mecanismos eficaces de reparación integral y sistemas de supervisión capaces de intervenir antes de que el daño alcance una difusión masiva.

Desde esta perspectiva, la regulación de la ciberviolencia de género exige abandonar una concepción centrada exclusivamente en la conducta individual del agresor para avanzar hacia un modelo que incorpore la responsabilidad de las plataformas, la transparencia de los sistemas algorítmicos y mecanismos sancionatorios proporcionales. Si el daño puede producirse a través de una arquitectura integrada de datificación biométrica, microsegmentación y amplificación algorítmica, la respuesta jurídica debe adoptar un enfoque igualmente sistémico, capaz de intervenir sobre cada uno de esos niveles sin sacrificar las garantías propias del Estado constitucional de Derecho.

Conclusiones

El análisis desarrollado permite sostener que la ciberviolencia de género facilitada por inteligencia artificial constituye un fenómeno estructural que no puede comprenderse desde las categorías tradicionales de la agresión

individual. La convergencia entre datificación biométrica, microsegmentación y amplificación algorítmica configura una arquitectura tecnológica que permite manipular y reproducir representaciones de la identidad de las mujeres y amplificar contenidos lesivos, produciendo efectos que trascienden el daño individual y alcanzan el ejercicio efectivo de la igualdad, la libertad de expresión y la participación democrática. Desde esta perspectiva, el principal aporte del estudio consiste en explicar cómo estas dimensiones no operan aisladamente, sino que integran una misma arquitectura del daño digital, cuya comprensión exige respuestas jurídicas igualmente sistémicas.

En el ámbito jurídico, el análisis comparado permite identificar la relevancia de los deberes de diligencia de las plataformas y de la transparencia algorítmica para prevenir y mitigar la amplificación de contenidos lesivos. La comparación entre el modelo europeo y el estadounidense evidencia diferencias en el alcance de las obligaciones impuestas a los intermediarios digitales, mientras que la jurisprudencia constitucional colombiana aporta elementos para comprender la transparencia algorítmica como garantía de los derechos fundamentales. Estos antecedentes respaldan la necesidad de articular mecanismos de responsabilidad, prevención y reparación que complementen la persecución penal de las conductas más graves.

La investigación no agota las problemáticas examinadas. En particular, resulta necesario avanzar hacia estudios empíricos que permitan medir el impacto de la autocensura y del retiro de las mujeres de los espacios digitales como consecuencia de la violencia algorítmicamente amplificada, así como profundizar el análisis comparado de mecanismos jurídicos que hagan exigibles los deberes de diligencia, transparencia y rendición de cuentas de las plataformas digitales. Ambas líneas permitirán evaluar con mayor precisión la eficacia de los modelos regulatorios emergentes y contribuir al diseño de respuestas compatibles con la protección de los derechos fundamentales en los entornos digitales.

Referencias

Ábrego Molina, Victor Hugo y Flores Mérida, Antony. 2021. Datificación crítica: práctica y producción de conocimiento a contracorriente de la gubernamentalidad algorítmica. Dos ejemplos en el caso mexicano. Revista Administración Pública y Sociedad, (11), 211-229. https://revistas.unc.edu.ar/index.php/APyS/article/view/33025/34263

Álvarez Buján , María Victoria. 2023. Inteligencia artificial y medidas cautelares en el proceso penal: tutela judicial efectiva y autodeterminación informativa en potencial riesgo. Revista Española de Derecho Constitucional, 127, 177-207. https://doi.org/10.18042/cepc/redc.127.06

Araujo Pulido, Fernando Javier, Gámez Chávez, Diana Cristina, y Morales Vargas, Jair de Jesús. 2026. Tecnoviolencias y violencia simbólica en entornos digitales. Transdigital, 7(13), e583. https://doi.org/10.56162/transdigital583

Arenas, Francisco. 2024. Más allá del capitalismo de la vigilancia. En E. González-Esteban y J. C. Siurna Aparisi (Eds.), Inteligencia Artificial: Concepto, alcances y retos (pp. 121-147). Tirant Humanidades.

Avendaño Paredes, Carlos Luis, Sarraute Requesens, María Magdalena, Torres Zambrano, Yorneylis, y Hidalgo López, Reinaldo. 2025. Fronteras digitales y violencia de género: Un estado del arte en América Latina. EDP University of Puerto Rico; Universidad Abierta Interamericana; Cátedra UNESCO "Universidad e Integración Regional" (UNAM). ISBN: 978-1-950792-99-3

Balkin, Jack. 2012. El futuro de la libertad de expresión en una era digital. Derecho y Humanidades, 20, 41-62. https://derechoyhumanidades.uchile.cl/index.php/RDH/article/view/34814

Barlow, Charlotte, y Awan, Imran. 2016. "You need to be sorted out with a knife": The attempted online silencing of women and people of Muslim faith within academia. Social Media + Society, 2(4). https://doi.org/10.1177/2056305116678896

Castillo Morales, Juan Pablo. 2019. Metodología y comparación jurídica en el Derecho penal. La incidencia del Derecho comparado en la estructura de la dogmática jurídico-penal. Revista de Derecho, (246), 13-47. http://dx.doi.org/10.4067/S0718-591X2019000200013

Celuch, Magdalena, Latikka, Rita, Oksa, Reetta, y Oksanen, Atte. 2023. Online harassment and hate among media professionals: Reactions to one's own and others' victimization. Journalism & Mass Communication Quarterly. https://doi.org/10.1177/10776990221148987

Centro de Derechos Humanos UDP. (2024). Informe anual sobre derechos humanos en Chile 2024. Ediciones Universidad Diego Portales. https://derechoshumanos.udp.cl/cms/wp-content/uploads/2024/11/INFORME-ANUAL-DDHH-UDP-2024-COMPLETO.pdf

Chesney, Bobby; Citron, Danielle. 2019. Deep Fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(1753). https://doi.org/10.15779/Z38RV0D15J

Corte Constitucional de Colombia. 2025. Upegui Mejía contra Agencia Nacional Digital AND, 26 de febrero de 2025, T-067/25.

Defensoría de la Niñez. 2024. Informe anual 2024: Cuenta pública y derechos humanos de niños, niñas y adolescentes en Chile. https://www.defensorianinez.cl/informe-anual-2024/

Domínguez Álvarez, José Luis. 2022. La silenciosa construcción de los derechos digitales. Revista Boliviana de Derecho, (34), 582-613. https://www.revista-rbd.com/numero-34/

Ferrante, Enzo. 2021. Inteligencia artificial y sesgos algorítmicos ¿Por qué debería importarnos? Nueva Sociedad, 294, 27-36. https://dialnet.unirioja.es/servlet/articulo?codigo=8027696

Finol Romero, Lorayne y Flor María Ávila Hernández. 2026. "Transparencia algorítmica como garantía de la dignidad humana: una propuesta convergente para la gobernanza de la IA en Chile y Colombia". Novum Jus 20, núm. 1 2026: 289-322. https://doi.org/10.14718/NovumJus.2026.20.1.10

García, María Fernanda. 2026. Violencia de género en entornos digitales. Tensiones a propósito de su criminalización. Revista de la Academia Neolonesa de Ciencias Penales, 2(3), 1-17. https://doi.org/10.12345/6pn2za64 . https://www.revistaancp.net/index.php/rancp/article/view/37

García-Estévez, Noelia, Ballesteros-Aguayo, Lucía, y Colussi, Juliana. 2025. Desinformación y manipulación de la opinión pública: una revisión sistemática sobre astroturfing (2004-2024). Revista De Comunicación, 24(2), 159-181. https://doi.org/10.26441/RC24.2-2025-3988

Gendler, Martín. 2024. Datificación social e inteligencia artificial: ¿hacia un nuevo "salto de escala"? Resonancias. Revista de Filosofía, (17), 121-141. https://doi.org/10.5354/0719-790X.2024.74503

Gobierno de España, Ministerio de la Presidencia, Justicia y Relaciones con las Cortes, Ministerio para la Transformación Digital y de la Función Pública, y Ministerio de Juventud e Infancia. 2026. Informe técnico sobre la posible responsabilidad penal derivada de la generación y difusión de contenidos sexuales y pornografía infantil mediante IA en redes sociales. Gobierno de España. https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/presidencia/Documents/2026/170226-informe-delitos-redes-sociales.pdf

G'sell, Florence. 2025. Digital Authoritarianism: from state control to algorithmic despotism. http://dx.doi.org/10.2139/ssrn.5117399

Jane, Emma. 2017. Misogyny online: A short (and brutish) history. Sage Publications. https://doi.org/10.4135/9781473916029

Kennedy, Helen, Poell, Thomas, y Van Dijck, José. 2015. Data and agency. Big Data & Society, 2(2), 1-7. https://doi.org/10.1177/2053951715621569

Lafont, Cristina. 2025. La democracia deliberativa tras la transformación digital de la esfera pública. Revista Latinoamericana sobre Democracia, (0), 20-29. https://doi.org/10.22201/iis.rld.2025.00.4

Lazard, Lisa, Capdevila, Rose, Turley, E. L., Gilfoyle, Kathryn, y Stavropoulou, Nelli. 2025. Deepfake technology and gender-based violence: A scoping review. Trauma, Violence, & Abuse (early access). https://doi.org/10.1177/15248380251384271

Ley N.º 21.719. Regula la protección y el tratamiento de los datos personales y crea la Agencia de Protección de Datos Personales. Diario Oficial de la República de Chile, 13 de diciembre de 2024.

Lucas, Rob. 2020. Capitalismo de vigilancia. Nueva Sociedad, (290), 46-58.

McGlynn, Clare, Rackley, Erika, y Houghton, Ruth. 2017. Beyond 'Revenge Porn': The Continuum of Image-Based Sexual Abuse. Feminist Legal Studies, 25, 25-46. https://doi.org/10.1007/s10691-017-9343-2

Milan, Stefania, y Beraldo, Davide. 2024. Data in movement: the social movement society in the age of datafication. Social Movement Studies, 23(3), 265-284. https://doi.org/10.1080/14742837.2024.2331550

Mota Orlob, Diego. 2023. Inteligencia artificial y comunicación política: "cuando la tecnología toma el poder". Friedrich-Ebert-Stiftung, FES Comunicación. https://biblioteca.registraduria.gov.co/opac_css/doc_num.php?explnum_id=156

Neuwirth, K., Frederick, E., y Mayo, C. 2007. The spiral of silence and fear of isolation. Journal of Communication, 57(3), 450-468. https://doi.org/10.1111/j.1460-2466.2007.00352.x

Noelle-Neumann, Elisabeth.1974. The spiral of silence: A theory of public opinion. Journal of Communication, 24(2), 43-51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Parlamento Europeo y Consejo de la Unión Europea. 2022. Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo, de 19 de octubre de 2022, relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales). Diario Oficial de la Unión Europea, L 277, 1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

Parlamento Europeo y Consejo de la Unión Europea. 2016. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de Datos). Diario Oficial de la Unión Europea, L 119, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Pauner Chulvi, Cristina. 2023. Transparencia algorítmica en los medios de comunicación y las plataformas digitales. Revista Española de la Transparencia, 17, 107-136. https://doi.org/10.51915/ret.308

Protection for private blocking and screening of offensive material, 47 U.S.C. § 230 1996. https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title47-section230

Ramos-Zaga, Fernando. 2026. La paradoja de seguridad y libertad en la era digital: un análisis crítico desde perspectivas biopolíticas y filosóficas. Revista de la Secretaría del Tribunal Permanente de Revisión, 14(23), e568. https://doi.org/10.16890/rstpr.a14.n23.e568

Rebollo Delgado, Lucrecio. 2023. Inteligencia artificial y derechos fundamentales. Dykinson S.L.

Reis, Ana Cláudia Borges Coutrim Dos; Campos, Francisco Itami. 2026. Inteligência Artificial no Judiciário Brasileiro: Sustentabilidade Digitalmente Auditável, Supervisão Humana, Legitimidade Democrática e Soberania Tecnológica. Fronteira: Journal of Social, Technological and Environmental Science, [S. l.], v. 15, n.

2, p. 184–204, 2026. DOI: 10.21664/2238-8869.2026v15i2.8415. Disponível em: https://periodicos.unievangelica.edu.br/index.php/fronteiras/article/view/8415

Ron Zambrano, Marlon y Rodríguez Jaramillo, Gabriela. 2025. El astroturfing potenciado por la Inteligencia Artificial: la manipulación de la opinión pública en la e-democracia de América Latina. Revista Especializada Justicia Electoral y Democracia, 8(9), 49-75. https://revista.tce.gob.ec/index.php/rjed/article/view/85

Rubio Núñez, Rafael; Franco Alvim, Frederico; de Andrade Monteiro, Vitor. 2024. Inteligencia artificial y campañas electorales algorítmicas. Disfunciones informativas y amenazas sistémicas de la nueva Comunicación política. Centro de Estudios Políticos y Constitucionales. https://www.cepc.gob.es/publicaciones/monografias/inteligencia-artificial-y-campanas-electorales-algoritmicas-disfunciones-informativas-y-amenazas

Sosa Huapaya, Alex. 2024. Conflictos entre la autodeterminación informativa y la segmentación de perfiles a través de la publicidad programática online en el Perú. Derecho & Sociedad, (63), 285-298. https://doi.org/10.18800/dys.202402.019

Tasente, Tanase, y Popescu, Cristina-Dana. 2025. La mayoría falsa: Dentro de la maquinaria invisible del astroturfing. Un modelo de mediación en serie. Revista ICONO 14: Revista Científica de Comunicación y Tecnologías Emergentes, 23(1), e2278. https://doi.org/10.7195/ri14.v23i1.2278

Torres-Zambrano, Yorneylis, y Morillo-Puente, Solbey. 2025. Violencia de género en línea. ¿Avanzando hacia la equidad digital? FEMERIS: Revista Multidisciplinar de Estudios de Género, 10(3), 54-82. https://doi.org/10.20318/femeris.2025.9744

Van Dijck, Jose. 2014. Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. Surveillance & Society, 12(2), 197-208. https://doi.org/10.24908/ss.v12i2.4776

Viña Castillo, Karla Brigida, Mantilla Burbano, Gabriela Aalexandra, y Quishpe Cherres, Lucía Elizabeth. 2025. "Golpes digitales": La nueva violencia contra la mujer en redes sociales, IA y ciberespacios. Polo del Conocimiento, 10(11), 2065-2093. https://doi.org/10.23857/pc.v10i11.10750

Zuboff, Shoshana. 2020. La era del capitalismo de la vigilancia: La lucha por un futuro humano frente a las nuevas fronteras del poder (A. Santos Mosquera, trad.) Paidós. https://perio.unlp.edu.ar/catedras/tpm/wp-content/uploads/sites/210/2023/06/ZUBOFF_1.pdf