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

Article

Datafication and Gender Violence: The Architecture of Silencing

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

[1] Doctor of Law. Universidad de Las Américas. ORCID: 0000-0003-4085-3000. Email: mrossel@udla.cl
[2] Doctor of Law. 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

Introduction

The contemporary digital revolution has not only transformed the forms and channels of communication, but also some of the structural conditions under which violence against women is exercised. Thus, artificial intelligence (AI) systems, recommendation algorithms and digital platforms have become central elements of public life in our times, mediating access to information, the formation of public opinion and social interaction. Regardless of the benefits that this new scenario brings with it, this same technological infrastructure -which

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

v.15, n.3, 2026 • p. 1-14. • 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.

does not constitute a neutral space- can reproduce and amplify pre-existing power asymmetries, among them the structural inequalities that have historically negatively affected women in the public space.

In this context, gender violence in digital environments can no longer be understood, analyzed and explained exclusively on the basis of the classical categories marked by an interpersonal aggression (physical, verbal, psychological, economic, etc.), carried out by and attributed to an identifiable aggressor. In this way, the interaction that occurs between the datafication of biometric characteristics, algorithmic microtargeting and artificial amplification strategies -for example, coordinated bot campaigns- can configure a digital infrastructure that facilitates the production of aggressions that are more personalized, permanent over time and scalable, making the legal attribution of responsibilities more difficult.

This new scenario poses important regulatory challenges: the opacity of algorithmic systems, the simultaneous intervention of natural persons and automated systems, in addition to the speed with which these forms of digital aggression evolve, strain the traditional categories of Law, which have been conceived and configured to attribute responsibilities on the basis of individual acts rather than infrastructures. In sum, limiting the State response to the criminal prosecution of individual aggressors is insufficient in this new scenario.

In this order of things, the present study postulates that biometric datafication and the algorithmic architecture of digital platforms together constitute an infrastructure that intensifies gender violence, conferring on it unprecedented capacities for reproduction, personalization and viralization of the harm caused to the victim. Furthermore, this new architecture not only individually victimizes the women exposed to these forms of violence, but also, in some way, fulfills a kind of "exemplary" function of collective scope: by naturalizing the cost of public exposure, it can favor self-censorship and the progressive withdrawal of women from spaces of democratic deliberation, thereby undermining goods essential to democracy and human dignity such as pluralism, equality and citizen participation.

In this way, the analysis of the relationships between technology, gender inequality and legal regulation also contributes to the debate on the social sustainability of digital environments, as spaces whose configuration must favor conditions of equality, inclusion and democratic participation.

The present study is structured as follows. After giving an account of the methodological approach employed, the transition from the datafication of behaviors toward the datafication of individual biometric traits is analyzed, a condition that makes possible the use of tools such as *deepfakes* and other forms of manipulation of a person's digital identity. Subsequently, the functioning of recommendation algorithms and artificial amplification strategies -such as microtargeting and *astroturfing*- in the configuration of digital gender violence is examined. Lastly, this work concludes with a discussion of the regulatory implications of this diagnosis, particularly regarding algorithmic transparency and the due diligence obligations of digital platforms.

Methodology

The present research adopts a legal-dogmatic methodology, aimed at identifying, interpreting and systematizing the legal norms, principles and categories relevant to the analysis of gender cyberviolence in digital environments mediated by artificial intelligence. From this perspective, the study examines the legal implications derived from biometric datafication, algorithmic microtargeting and the automated amplification of content, with the purpose of evaluating the sufficiency of the regulatory frameworks in force for the protection of women's fundamental rights.

Nevertheless, the phenomenon analyzed goes beyond the strictly normative sphere. As Castillo Morales (2019, 16-19) maintains, legal dogmatics acquires greater explanatory capacity when it incorporates tools from other disciplines that make it possible to understand the social and technological processes on which Law operates. Consequently, the research adopts an interdisciplinary approach and resorts to comparative law, integrating contributions from constitutional law, criminal law and data protection law, together with contributions from studies on digital governance, artificial intelligence, political communication and digital platforms.

The legal comparison focuses on the regulatory developments of the European Union and the United States, incorporating specific references to Latin American experiences, particularly in matters of digital platform liability, algorithmic transparency, protection of biometric data and digital violence against women. This comparative analysis of a selective nature makes it possible to identify differences between the regulatory approaches examined, evaluate their main strengths and limitations and formulate legal criteria applicable to the protection of fundamental rights in the face of the new forms of violence facilitated by artificial intelligence.

The primary sources analyzed comprise international human rights instruments, national and comparative legislation, regulatory documents on artificial intelligence and data protection, as well as a relevant ruling of the Constitutional Court of Colombia. The secondary sources correspond to specialized scientific literature, reports

prepared by international organizations and recent academic studies on datafication, algorithms, hate speech and digital gender violence, examined through a narrative literature review. The treatment of the sources was carried out through dogmatic analysis and systematic interpretation, which made it possible to construct the analytical categories used to explain the relationship between biometric datafication, algorithmic amplification and digital gender violence.

This methodological approach makes it possible to respond to the objectives of the research: to explain the existing relationship between biometric datafication, algorithmic architecture and gender cyberviolence, to critically evaluate the regulatory models examined and to propose legal criteria aimed at strengthening the protection of women's fundamental rights in digital environments.

Results and Discussion

Datafication of individual traits: from behavioral data to biometric data

The literature on social datafication has historically placed the emphasis on behavioral data, understood as the recording and quantification of preferences, interactions and consumption patterns that make it possible to segment audiences for commercial, political or other purposes. Along these lines, Kennedy et al. (2015, 1) define datafication as the process by which aspects of the social world not previously quantified are converted into data, comprising both demographic or profiling information and behavioral metadata -geolocation, interaction patterns or usage timestamps, among others-. This conception has served to explain how the massive collection of data makes it possible to know, predict and influence human behavior (Arenas 2024, 121-124; Ábrego Molina and Flores Mérida 2021, 214-216).

The progressive incorporation of biometric data into these datafication processes raises, however, an additional dimension: the use of physical and physiological characteristics as elements of identification and representation of persons. This attribution of privileged identificatory value to biometric data is not accidental, but responds to what Van Dijck (2014, 201) calls "dataism": the belief in the capacity of data to objectively represent social reality, which allows aggregated data -or, in this case, individualized biophysical traits- to be interpreted as direct and objective access to the truth about a subject. Transferred to the biometric level, this "dataist" logic helps explain why the face or voice synthesized through generative AI are socially plausible as an authentic representation of the victim: dataism naturalizes the idea that biophysical data, once extracted and algorithmically recomposed, continue to be a faithful correlate of people's real identity, even though the content produced is entirely false (Chesney and Citron 2019, 1757).

From a broader perspective, Gendler (2024, 122-124) situates these processes within a genealogy of biopolitical power, supported by developments in medicine, biology and political economy, in which data have been used to model "subjectivities, bodies, buildings and the management of actions and behaviors". However, both the classical definition of Kennedy *et al.* and the historical characterization proposed by Gendler were conceived mainly to explain the datafication of individual and collective behavior, although they do not exhaust the specific implications of the contemporary use of biometric traits. That framework is today insufficient for understanding a qualitatively different transformation: one that no longer falls on what the subject does, but also on what the subject is. Indeed, the progressive capture and processing of the face, the voice and other biometric traits shifts the object of datafication from behavior toward the very biophysical identity of persons.

From a legal perspective, this distinction is not merely descriptive, but has first-order normative consequences. Under the GDPR, special-category protection covers biometric data processed for the purpose of uniquely identifying a natural person**.** Personal data protection frameworks -from the European General Data Protection Regulation to the recent Law 21.719 in Chile- recognize biometric data as a sensitive or special category, subject to a regime of reinforced protection precisely because their unauthorized processing compromises not only informational privacy, but also rights linked to the identity of the data subject (Regulation [EU] 2016/679, art. 9; Law No. 21.719, 2024). For this reason, Álvarez Buján (2023, 185-188) warns that the use of artificial intelligence systems requires reinforcing the guarantees of effective judicial protection and informational self-determination, especially when automated decisions affect fundamental rights (Rebollo Delgado 2023, 122).

The relevance of this transformation goes beyond, however, the sphere of personal data protection. The permanent datafication of biometric traits modifies the traditional conditions of anonymity in the public space. As Ramos-Zaga (2026, 14) maintains, facial recognition technologies make it possible to link physical presence with digital identities and personal databases, weakening the expectation of anonymity associated with transitory presence in public spaces. In this context, biometrization can turn physical identity into an object of continuous recording, classification and technological reuse.

This second form of datafication —that of the individual trait rather than that of aggregate behavior— constitutes one of the technological preconditions that can facilitate the production of so-called *deepfakes*: hyperrealistic audiovisual content generated by training machine learning models on banks of images, videos or voice recordings of a specific person. While microtargeting seeks to identify whom to persuade, biometric datafication makes it possible to decide about whom to construct the message. The object of algorithmic intervention ceases to be only the receiver and also becomes the digital identity of the victim. Unlike behavioral microtargeting, which operates on the audience that receives a message, biometric datafication operates on the subject represented, transforming their image and voice into a resource that can be manipulated and reproduced by third parties.

This evolution can be read, moreover, as a manifestation of so-called surveillance capitalism. In the same way that behavioral data is transformed into a "surveillance asset" and into raw material for the production of marketable "behavioral futures" (Lucas 2020, 51), the face, the voice and other biometric traits acquire a new economic value as resources susceptible to extraction, storage and reuse. Biometric datafication can thus extend the extractive logic characteristic of surveillance capitalism from behavior toward the physical identity of persons. In the context of digital gender violence, however, that instrumentalization no longer pursues only commercial purposes, but also the deliberate production of harm through the manipulation of the victims' digital identity, conditioning the behavior of individuals and restricting the effective exercise of fundamental rights (G'sell 2025, 8-12).

The transition from behavioral datafication toward biometric datafication also entails a transformation of the very object of digital power. While the former seeks to predict and modify future behaviors through probabilistic inferences about individual preferences, the latter makes it possible to intervene directly on the constitutive attributes of personal identity. The face, the voice or gestures cease to be only identifying elements and become inputs susceptible to extraction, reproduction and manipulation through artificial intelligence. Consequently, the physical traits used in processes of digital identification or representation can serve not only to identify a person, but also to generate synthetic representations of them.

This transformation has a normative consequence that Ramos-Zaga (2026, 20) formulates as a public policy criterion: the need to enshrine "the inalienability of the biometric footprint, elevating anonymity in public spaces to the category of a non-negotiable fundamental right". This proposal of inalienability is fully coherent with the legal distinction between behavioral data and biometric data: if the former admits regimes of consent and portability -characteristic of the contractual logic of personal data-, the latter, because it constitutes the very biophysical identity of the subject, requires a status of reinforced protection which, according to said proposal, cannot be subject to the contractual disposal of the data subject, precisely because its non-consensual appropriation (as can occur with the production of *deepfakes*) violates not only informational privacy, but the integrity of identity itself.

Biometrization does not by itself create new forms of misogyny or of violence against women. It provides, however, a technological infrastructure capable of intensifying historically existing forms of violence through unprecedented mechanisms of reproduction, automation and viralization of harm. In digital environments, platforms can not only amplify pre-existing structural inequalities, but they facilitate new modalities of dissemination and persistence of aggressions, which can increase their reach and permanence. In this sense, biometric datafication constitutes a technical precondition that enhances traditional forms of gender violence, endowing them with capacities for manipulation, scalability and circulation that acquire new dimensions in digital environments (Avendaño Paredes *et al.* 2025, 6-8; Viña Castillo *et al.* 2025, 2069-2070).

This technological mutation has special relevance for digital gender-based violence. If classical microtargeting seeks to identify the audiences most susceptible to receiving certain messages, biometric datafication shifts the focus toward the victim herself, whose digitized body comes to constitute the object of manipulation. The extraction and reuse of images, voices and other personal traits makes it possible to produce highly plausible false representations -as occurs with sexual deepfakes- which are subsequently inserted into algorithmic ecosystems capable of amplifying hate speech, harassment campaigns and coordinated violence.

In short, biometric datafication transforms the physical identity of persons into an object susceptible to extraction, classification and algorithmic manipulation. This shift from behavioral data toward biometric data not only redefines the scope of the legal protection of personal data, but contributes to configuring the technological infrastructure on which microtargeting, recommendation systems and the amplification strategies that characterize contemporary digital violence subsequently operate. Understanding this transformation is indispensable for analyzing how algorithmic architecture converts that raw material into mechanisms for the production, distribution and escalation of harm against women, an issue examined in the following section.

Algorithms, Platforms and Hate Speech: Gender Cyberviolence in the Digital Ecosystem

The analysis developed in the preceding section made it possible to explain how the incorporation of biometric datafication into behavioral datafication processes has transformed the identity of persons into an object susceptible to extraction, classification and algorithmic manipulation. Nevertheless, the explanation of gender cyberviolence is not exhausted by the obtaining and reuse of biometric data. It is equally necessary to analyze the technological infrastructure that determines the circulation, visibility and amplification of those data and of the content constructed from them. This requires shifting the analysis from the logic of datafication toward the logic of the algorithmic distribution of attention (Rubio *et al.* 2024, 31-34).

Digital platforms and their recommendation systems constitute the core of that second dimension of the problem. If biometric datafication provides one of the sources of information on which synthetic content can be constructed, it is the algorithmic architecture that influences which content acquires visibility, to which audiences it is directed and with what intensity it is amplified (Mota 2023, 14-18). It is precisely in the interaction between biometric datafication, microtargeting and recommendation systems that the contemporary architecture of gender cyberviolence is configured and where hate speech against women ceases to depend exclusively on the action of an individual aggressor and becomes a phenomenon of structural dissemination and amplification.

The digital revolution promised a space of horizontal communication and democratization of knowledge; however, cyberspace has proven to be far from constituting a neutral environment. As Avendaño Paredes *et al.* (2025, 14) warn, the digital ecosystem reproduces and amplifies the power asymmetries and structural forms of violence present in the offline world, consolidating patriarchal dynamics that harass, silence and subordinate women. In this context, the intersection between algorithmic profiling, microtargeting and the automated propagation of hate speech has configured a scenario of special vulnerability that poses unprecedented challenges for Law and for criminal policy.

Digital platforms have ceased to play a merely instrumental role and have become infrastructures that organize the circulation of information and condition the formation of public opinion. Within the framework of so-called surveillance capitalism, the systematic extraction of behavioral and biometric data constitutes the economic foundation of business models based on the prediction and modulation of human behavior (Ramos-Zaga 2026, 11-13). Consequently, recommendation algorithms not only determine what content each user views, but can also condition the intensity of their exposure, shape preferences and favor processes of polarization (Araujo Pulido *et al.* 2026, 11; Ramos-Zaga 2026, 5; Viña Castillo *et al.* 2025, 2069-2070).

This logic of prioritization does not constitute an accidental consequence of technological design, but a manifestation of the economic model that structures the functioning of digital platforms. Within the framework of the theory of surveillance capitalism developed by Zuboff, so-called "instrumentarianism" entails the systematic use of data and human behavior for purposes of prediction, modification, monetization and control (Lucas 2020, 51). In this context, people's attention ceases to be a mere cognitive resource and becomes a marketable asset, so that algorithms tend to privilege content capable of generating higher levels of interaction and time spent on the platforms. Emotionally intense content -among it hate speech, misogyny and other forms of violence directed against women- can be especially functional to that logic, insofar as it increases the production of behavioral data and strengthens the predictive capacity of algorithmic systems. Consequently, the amplification of this type of content does not constitute a collateral effect that can be corrected through isolated technical adjustments, but must also be examined in relation to the business model that sustains the surveillance economy and the contemporary recommendation architecture.

In this scenario, microtargeting has ceased to be a marketing tool intended to personalize commercial messages. It currently constitutes one of the main mechanisms through which platforms determine which content will be shown, to which audiences and with what degree of intensity it will be disseminated. In this way, microtargeting ceases to be solely a technique of personalized advertising and becomes an infrastructure for the differential distribution of digital harm. The recommendation algorithms of platforms such as Instagram, X or TikTok can favor the viralization of highly emotional, controversial or degrading content, generating feedback loops that intensify sexualization, misogyny and other forms of digital violence (Celuch *et al.* 2023, 625; Gobierno de España, 2026, 6). Added to this is that artificial intelligence systems, by being trained on data that reflect historical inequalities, can reproduce structural biases that perpetuate discriminatory stereotypes and deepen phenomena of algorithmic discrimination (Ferrante 2021, 27-36; Centro de Derechos Humanos UDP 2024, 368).

Consequently, gender violence in the digital environment can no longer be understood exclusively as the result of the conduct of an individual aggressor. It is inserted, rather, into a genuine architecture of digital harm, in which human aggressors, technological platforms, algorithmic systems and economic models sustained by

the attention economy and surveillance capitalism converge (Ramos-Zaga 2026, 24; Viña Castillo *et al.* 2025, 2070). Datafication constitutes one of the technical preconditions of that architecture, since it converts human activity —and, as explained in the preceding section, also the biometric traits that individualize persons— into data susceptible to classification, prediction and algorithmic reuse. In this context, algorithms cease to be neutral instruments and operate as true sociotechnical devices that distribute public attention, rank the visibility of certain content and can amplify dynamics of confrontation and harassment (Araujo Pulido *et al.* 2026, 11).

Through microtargeting and algorithmic profiling, mass processing systems use *proxies* —apparently neutral variables, such as postal codes, consumption patterns or browsing habits— that, in practice, maintain a close correlation with protected categories such as gender, race or social class (Ramos-Zaga 2026, 24). This can produce a disparate impact (*disparate impact*), reproducing and amplifying pre-existing structural inequalities through automated decisions that disproportionately affect certain groups (Ramos-Zaga 2026, 24 and 25; Sosa Huapaya 2024, 288-290). In turn, being designed to maximize *engagement*, algorithms privilege controversial, violent or sexualized content, increasing women's exposure to dynamics of harassment and giving rise to what part of the doctrine identifies as algorithmic violence (Araujo Pulido *et al.* 2026, 11; Viña Castillo *et al.* 2025, 2069-2070). Coordinated manipulation strategies such as astroturfing are also articulated on this technological infrastructure, through which bots, fake accounts and networks of *trolls* artificially fabricate the appearance of social support or rejection in order to amplify harassment campaigns and hate speech. Covert coordination constitutes precisely the distinctive element of these practices, since it seeks to present as a spontaneous expression of citizens what in reality corresponds to deliberately organized actions (Tasente and Popescu 2025, 6-8; García-Estévez *et al.* 2025, 173-175).

Data infrastructure and algorithmic systems do not have, by themselves, an intrinsically antidemocratic meaning. As Milan and Beraldo (2024, 272) show, these same tools can be reappropriated by civil society to favor processes of democratic participation and collective action. Consequently, the difference between a data infrastructure oriented toward strengthening public deliberation and another intended to amplify hate speech does not lie exclusively in the technology employed, but also in the purposes that guide its use and in the power dynamics it reproduces. From this perspective, algorithmic architecture constitutes an ambivalent infrastructure, capable both of favoring collective agency and of intensifying forms of manipulation and digital violence. It is precisely this second dimension that will be analyzed in the following section, through the examination of the main manifestations of gender cyberviolence enhanced by algorithmic microtargeting.

Manifestations of Violence Enhanced by Microtargeting

The algorithmic architecture described above not only conditions the circulation of content, but can also transform the modalities through which gender violence is exercised in digital environments. The combination of microtargeting, artificial intelligence and automated recommendation systems allows aggressions to be increasingly personalized, persistent and scalable, especially affecting women with high public exposure, such as journalists, activists or political representatives (Araujo Pulido *et al.* 2026, 4-11; Viña Castillo *et al.* 2025, 2071).

These transformations manifest themselves mainly in three interrelated mechanisms through which algorithmic architecture intensifies gender violence in digital environments: (i) the automation of aggressions; (ii) the personalization of harm through microtargeting; and (iii) the manipulation of the information ecosystem through mechanisms of artificial amplification and false consensus.

First, the automation of aggressions constitutes one of the main manifestations of contemporary digital violence. Generative artificial intelligence has facilitated the creation of non-consensual sexual *deepfakes*, whose dissemination can be massively replicated through automated systems, considerably expanding the reach of the harm and making it difficult to identify those responsible (Viña Castillo et al. 2025, 2070). Added to this are coordinated harassment campaigns carried out through *troll* farms (*troll centers*) and automated accounts (*bots*), which generate waves of misogynistic insults, transphobic speech and threats directed against women with high public exposure, such as journalists, activists and political representatives, with the purpose of discrediting their work and forcing their withdrawal from public debate (Araujo Pulido *et al.* 2026, 4; Avendaño Paredes *et al.* 2025, 32). The combined intervention of human actors and automated systems can make it difficult to identify individual conducts and attribute responsibilities, configuring a decentralized, persistent violence that is particularly difficult to investigate from the legal perspective (Viña Castillo *et al.* 2025, 2075).

Second, the personalization of harm makes it possible to direct specific attacks against women previously profiled through microtargeting techniques and mass data processing. On the basis of information obtained from the victims' digital activity, platforms and aggressors can identify particularly vulnerable profiles and design targeted harassment campaigns. In this context, practices such as *doxxing* —the unauthorized disclosure of personal data—, the dissemination of false or decontextualized information and smear campaigns are aimed at

affecting the professional reputation, public credibility and personal safety of the victims (García 2026, 4; Avendaño Paredes *et al.* 2025, 27). These aggressions usually focus on gender stereotypes, questioning women's physical appearance, sexuality or professional capacity as mechanisms of silencing and exclusion from the public space (Avendaño Paredes *et al.* 2025, 20).

Third, the manipulation of the information ecosystem amplifies and legitimizes violence through coordinated strategies that simulate non-existent social consensus. Among them, astroturfing stands out, a manipulation technique that uses fake accounts, *sock puppets* and bot networks to feign spontaneous citizen support when, in reality, it responds to organized disinformation campaigns (García-Estévez *et al.* 2025, 161). Through these mechanisms, hate speech and anti-feminist narratives are disseminated in a coordinated manner, creating the false perception that there is a broad and socially shared rejection of demands for equality or of certain women because of their gender (García-Estévez *et al.* 2025, 162). This appearance of consensus not only amplifies the reach of violent messages, but can also contribute to normalizing misogyny, reinforcing processes of polarization and eroding the conditions for democratic deliberation based on authentic information.

These three mechanisms —the automation of aggressions, the personalization of harm and the manipulation of the information ecosystem— do not operate in isolation, but feed back into one another. Biometric datafication can provide the information used to construct synthetic content; microtargeting makes it possible to identify potential audiences and victims; automated systems amplify its circulation; and coordinated manipulation campaigns artificially reinforce its legitimacy and reach. The result is a digital infrastructure capable of producing persistent, scalable harms that are difficult to attribute legally, which poses additional challenges to traditional models of liability for interpersonal violence.

In this way, the manifestations described do not only produce individual effects on the victims, but project their effects onto the very functioning of the digital public space. When *deepfakes,* coordinated harassment, *doxxing* or *astroturfing* campaigns seek to discredit, intimidate or expel women from public debate, they not only harm rights such as dignity, equality, personal integrity or informational self-determination, but also alter the conditions under which public opinion is formed in digital environments (Reis and Campos 2026, 189). In this sense, the Spanish Charter of Digital Rights expressly recognizes rights of a political nature, among them the right to freely receive information and the right to political participation through digital means (Domínguez Álvarez 2022, 604-605). The artificial manipulation of consensus through astroturfing strategies directed against women directly compromises these legal interests, by distorting the processes of public deliberation, weakening democratic pluralism and favoring the silencing of authentic voices. From this perspective, gender cyberviolence transcends the sphere of individual harm to also become a structural problem of democratic quality, insofar as it conditions who can participate in public debate, under what conditions they do so and which discourses achieve visibility within the digital ecosystem (Ron Zambrano and Rodríguez Jaramillo 2025, 49-75).

The psychosocial consequences of this architecture of violence are profound and persistent. The women affected may experience anxiety, isolation, effects on their identity and damage to their personal and professional reputation, which can favor processes of self-censorship and, in certain cases, their progressive exclusion from spaces of public deliberation and democratic participation (Araujo Pulido *et al.* 2026, 11; Viña Castillo *et al.* 2025, 2077; Lafont 2025, 22-26; Lazard *et al.* 2025, 10).

From the legal perspective, this technological architecture reveals important regulatory challenges. The opacity of algorithms —the so-called black boxes—, the simultaneous intervention of human actors and automated systems, as well as the speed with which these forms of aggression evolve, make it difficult to attribute responsibility and to adopt effective responses (Araujo Pulido *et al.* 2026, 11). In this context, limiting State intervention to the criminal prosecution of individual aggressors is insufficient. An adequate regulatory response requires incorporating due diligence obligations for digital platforms, strengthening algorithmic transparency standards and establishing effective mechanisms for the prevention, detection and mitigation of those recommendation systems that contribute to amplifying and perpetuating online gender violence (Torres-Zambrano and Morillo-Puente 2025, 58; Viña Castillo *et al.* 2025, 2070).

Silencing and Self-Censorship as a Disciplining Effect

The mechanisms analyzed in the preceding section —the automation of aggressions, the personalization of harm and the manipulation of the information ecosystem— do not constitute an end in themselves, but can converge in a common result: altering the conditions of women's participation in the digital public space (Barlow and Awan 2016, 2 and 3; Noelle-Neumann1974, 43-51; Neuwirth *et al.* 2007, 450-468). From this perspective, silencing and self-censorship cannot be understood solely as collateral effects of algorithmic architecture, but also as a structural consequence of the operating model of digital platforms. In Zuboff's (2020,

438) terms, this dynamic can be interpreted on the basis of instrumentarianism, understood as the systematic use of data and human behavior for purposes of prediction, modification, monetization and control. When digital aggressions induce women to withdraw from public debate, limit their participation or abandon certain digital spaces, violence ceases to operate solely as an individual aggression and becomes a mechanism for disciplining the public space (Araujo Pulido *et al.* 2026, 11). In this context, fear ceases to be solely a subjective reaction to violence and is transformed into a structural mechanism for regulating expression and democratic participation (Araujo Pulido *et al.* 2026, 11).

In this way, algorithmic architecture can not only amplify harmful content, but can also modify the conditions under which women exercise their rights of expression, participation and influence in digital environments. Silencing and self-censorship thus cease to be exclusively individual responses to violence and become indicators of the structural impact that recommendation and amplification technologies can exert on the effective exercise of fundamental rights in the digital public space.

The first manifestation of this disciplining effect is self-censorship. Faced with the risk of suffering harassment campaigns, public humiliation or threats of physical and sexual violence, many women voluntarily modify their digital behavior. This translates into a reduction in their posts, the limitation of their interactions, the abandonment of certain topics considered conflictive or the partial closure of their social media profiles (Avendaño Paredes *et al.* 2025, 37). Self-censorship thus ceases to be a fully free decision and becomes an adaptive response to a hostile communicative environment.

This characterization requires, however, a conceptual clarification. Self-censorship cannot be interpreted as the result of a fully free decision, nor as the simple negation of the victims' capacity for agency. As Kennedy *et al.* (2015, 2) point out, the relationship between structure and agency has a dialectical character: structures condition subjects' possibilities for action, but subjects continue to act within them and even to resist them.

Applied to gender cyberviolence, this perspective makes it possible to understand that self-censorship simultaneously constitutes an individual response to violence and the result of an algorithmic environment that progressively restricts the conditions for safe and equal participation in the digital public space. Consequently, silence does not represent a fully autonomous choice, but a form of agency exercised under structural conditions that significantly reduce the range of options available to women. From this perspective, self-censorship ceases to be a merely individual reaction and becomes an indicator of the impact that algorithmic architecture exerts on the effective exercise of freedom of expression, democratic participation and women's right to intervene on equal terms in digital environments (McGlynn *et al.* 2017, 26-28; Jane 2017, 13).

Self-censorship can lead to a progressive withdrawal from the digital public sphere. Continued exposure to campaigns of violence and discrediting can cause some women to abandon digital platforms or significantly reduce their participation in them, with direct consequences for their professional, academic and political opportunities (Avendaño Paredes *et al.* 2025, 38). This phenomenon acquires special seriousness in the political sphere, where digital harassment directed against female candidates, journalists or human rights defenders restricts the plurality of democratic debate and contributes to reproducing structural relations of gender inequality (Araujo Pulido *et al.* 2026, 11).

The withdrawal of women from the digital public space also admits a reading from the theory of datafied social movements. Milan and Beraldo (2024, 277) maintain that certain social actors may choose to partially withdraw from the dynamics of datafication as a strategy of resistance against surveillance mechanisms. Nevertheless, the authors themselves warn that such withdrawal does not imply escaping the logic of datafication, but rather reveals the reach of the structures that sustain it. Applied to gender cyberviolence, this reflection makes it possible to understand that the abandonment of certain digital spaces by women does not constitute a true exit from the algorithmic architecture, but a manifestation of its capacity to condition the forms of participation in the public space.

This disciplining effect is reinforced by the social normalization of digital violence. Frequently, aggressions are minimized under the idea that "they only happen on the internet" or justified through an absolute understanding of freedom of expression (Torres-Zambrano and Morillo-Puente 2025, 58). However, the repetition and viralization of hate speech can favor a process of social legitimation that turns violence into an apparently inevitable cost for women who actively participate in digital spaces (Avendaño Paredes *et al.* 2025, 14).

Consequently, gender trolling and hate speech do not constitute mere communicative excesses characteristic of social networks, but can operate as mechanisms of social control that seek to discipline women's participation through the production of fear, self-censorship and exclusion from public debate. From this perspective, digital violence not only violates individual rights, but compromises collective goods essential

to a democratic society, such as pluralism, equality in political participation and freedom of expression exercised under conditions of safety (García 2026, 5; Araujo Pulido *et al.* 2026, 11).

The analysis developed makes it possible to reconstruct an integrated architecture of production and amplification of gender cyberviolence. In it, biometric datafication can provide information susceptible to manipulation (1); the algorithmic infrastructure of platforms determines its circulation, differential distribution and amplification (2); these processes materialize in concrete mechanisms of automated, personalized and coordinated violence (3); and, finally, they can produce a structural effect of silencing and self-censorship that alters the conditions of women's participation in the digital public space (4).

This architecture not only compromises the fundamental rights of the victims considered individually, but also affects the structural conditions of democratic deliberation. By restricting the effective exercise of freedom of expression, equality and the political participation of women in digital environments, gender cyberviolence affects pluralism and the quality of public debate. Consequently, its legal treatment cannot be exhausted by the traditional mechanisms of personal data protection or individual liability, but requires a regulatory response that jointly considers biometric datafication, algorithmic architecture and the role played by digital platforms in the production, amplification and persistence of harm. The response of Law must therefore be oriented not only toward repairing the harm once produced, but also toward intervening on the structural conditions that allow its production and reproduction in digital ecosystems.

Regulatory challenges in the face of the algorithmic architecture of gender cyberviolence

The preceding analysis demonstrates that gender cyberviolence enhanced by artificial intelligence constitutes a structural phenomenon whose explanation goes beyond the traditional categories of criminal law or personal data protection. The interaction between biometric datafication, algorithmic architecture and recommendation systems reveals that harm is not produced solely by the individual action of the aggressor, but also by the functioning of platforms that can contribute to amplifying, distributing and prolonging the circulation of harmful content. Consequently, an effective regulatory response must intervene at each of these levels, combining mechanisms of liability, transparency and sanction, without disregarding the requirements derived from freedom of expression and other fundamental rights.

Liability of platforms and intermediaries

A first regulatory axis consists of redefining the legal regime of liability of digital platforms. The exclusive attribution of liability to the material author of the content is insufficient in the face of phenomena characterized by anonymity, automation and the rapid propagation of content through algorithmic systems. As Balkin (2012, 50) warns, in many cases it is considerably more effective to direct actions against the intermediaries that allow the dissemination of illicit content than to individually prosecute those who generate it.

In comparative law, two regulatory models can be identified. The first corresponds to the European model, consolidated through the Digital Services Act (*Digital Services Act,* European Union 2022), which imposes differentiated due diligence obligations on platforms, notice and takedown mechanisms for content, complaint procedures for users and audit obligations with respect to their algorithmic systems. The second corresponds to the North American model, built on Section 230 of the *Communications Decency Act* of 1996, which privileges innovation and self-regulation by excluding, as a general rule, the liability of platforms for content generated by third parties.

From the perspective developed in this work, the European model offers greater tools for confronting gender cyberviolence, precisely because it establishes obligations for platforms that go beyond liability for third-party content and extend to certain risks associated with the functioning of their services and recommendation systems.

Algorithmic transparency

A second regulatory axis consists of strengthening the mechanisms of algorithmic transparency. If the architecture of digital harm depends on the functioning of algorithms that select, rank and recommend content, their functioning cannot remain completely opaque to public control.

As Pauner Chulvi (2023, 123) maintains, algorithmic transparency fulfills a dual function. From an external perspective, it makes it possible to control the existence of errors, discrimination and algorithmic biases, as well as to guarantee the exercise of the rights of access and information. From an internal dimension, it facilitates the technical review of automated systems by programmers, analysts and independent auditors.

This understanding has also been adopted by the Constitutional Court of Colombia, in holding that algorithmic transparency constitutes an indispensable guarantee in a democratic society, since only through the availability of information is it possible to control the impact of automated systems on the autonomy, equality and dignity of persons (Corte Constitucional de Colombia 2025, Finol Romero and Ávila Hernández 2026, 310).

An integral regulatory model

Finally, the legal response cannot rest exclusively on the expansion of criminal law. Criminal sanction continues to be necessary for the most serious manifestations of digital violence, but it must operate in accordance with the principles of legality, proportionality and ultima ratio. In parallel, it is indispensable to strengthen the responses from administrative and civil law, through reinforced due diligence obligations for platforms, effective mechanisms of integral reparation and supervision systems capable of intervening before the harm reaches massive dissemination.

From this perspective, the regulation of gender cyberviolence requires abandoning a conception centered exclusively on the individual conduct of the aggressor in order to advance toward a model that incorporates the liability of platforms, the transparency of algorithmic systems and proportional sanctioning mechanisms. If harm can be produced through an integrated architecture of biometric datafication, microtargeting and algorithmic amplification, the legal response must adopt an equally systemic approach, capable of intervening at each of those levels without sacrificing the guarantees characteristic of the constitutional State governed by the rule of law.

Conclusions

The analysis developed makes it possible to maintain that gender cyberviolence facilitated by artificial intelligence constitutes a structural phenomenon that cannot be understood from the traditional categories of individual aggression. The convergence between biometric datafication, microtargeting and algorithmic amplification configures a technological architecture that makes it possible to manipulate and reproduce representations of women's identity and to amplify harmful content, producing effects that transcend individual harm and reach the effective exercise of equality, freedom of expression and democratic participation. From this perspective, the main contribution of the study consists of explaining how these dimensions do not operate in isolation, but form part of one and the same architecture of digital harm, whose understanding requires equally systemic legal responses.

In the legal sphere, the comparative analysis makes it possible to identify the relevance of the due diligence obligations of platforms and of algorithmic transparency for preventing and mitigating the amplification of harmful content. The comparison between the European and the United States models reveals differences in the scope of the obligations imposed on digital intermediaries, while Colombian constitutional case law provides elements for understanding algorithmic transparency as a guarantee of fundamental rights. These antecedents support the need to articulate mechanisms of liability, prevention and reparation that complement the criminal prosecution of the most serious conducts.

The research does not exhaust the problems examined. In particular, it is necessary to advance toward empirical studies that make it possible to measure the impact of self-censorship and of the withdrawal of women from digital spaces as a consequence of algorithmically amplified violence, as well as to deepen the comparative analysis of legal mechanisms that make enforceable the duties of diligence, transparency and accountability of digital platforms. Both lines will make it possible to evaluate more precisely the effectiveness of emerging regulatory models and to contribute to the design of responses compatible with the protection of fundamental rights in digital environments.

References

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