



 Cristian Alberto López Rodríguez *

Universidad Michoacana de San Nicolás de Hidalgo,
México
alberto_loprod@outlook.com

Cuadernos de Literatura

núm. 30, e3007, 2026
Universidad Nacional del Nordeste, Argentina
ISSN: 0326-5102
ISSN-E: 2684-0499
Periodicidad: Semestral
revistas@unne.edu.ar

Recepción: 15/03/2026
Aprobación: 20/07/2026

DOI: <https://doi.org/10.30972/clt.309673>

URL: <https://portal.amelica.org/ameli/journal/785/7855782009/>

Resumen: En el ecosistema digital, la “economía de la atención” y los algoritmos priorizan el *engagement* y la polarización mediante filtros burbuja y cámaras de eco. Mediante un enfoque cualitativo que integra un análisis estrictamente documental y jurídico-comparativo, este artículo examina la insuficiencia regulatoria frente a la violencia digital. Los resultados demuestran que el “pánico moral” actúa como un mecanismo discursivo encubierto, permitiendo que el discurso de odio se disfraze de preocupación legítima para eludir la moderación corporativa y vulnerar la dignidad de grupos históricamente marginados, particularmente mujeres y colectivos LGTBIQ+. Se concluye la inviabilidad de la autorregulación corporativa y se exige la intervención activa del Estado mexicano. Bajo el principio interamericano de que los derechos *offline* aplican *online*, es urgente un marco normativo que subordine el modelo de negocio algorítmico al respeto irrestricto de la dignidad humana.

Palabras clave: algoritmos, economía de la atención, discurso de odio, pánico moral, derechos humanos.

Abstract: In the digital ecosystem, the “attention economy” and algorithms prioritize engagement and polarization through filter bubbles and echo chambers. Through a qualitative approach that integrates a strictly documentary and

Notas de autor

- * Cristian Alberto López Rodríguez es Licenciado en Derecho por la Universidad Autónoma de Ciudad Juárez (UACJ) y Maestro en Derecho Procesal Constitucional por la Universidad Michoacana de San Nicolás de Hidalgo (UMSNH), donde fue galardonado con la Medalla “Ignacio Chávez Sánchez” al mérito académico. Cuenta con la distinción de Candidato a Investigador Estatal en el Padrón de Investigadoras e Investigadores de Michoacán (PIIM). Sus líneas de investigación se centran en el Sistema Interamericano de Derechos Humanos, el cumplimiento de sentencias internacionales, la perspectiva de género y la protección de grupos en situación de vulnerabilidad. Ha sido ponente en diversos congresos internacionales y es autor de capítulos de libro editados por Tirant lo Blanch, así como de múltiples artículos en revistas científicas indexadas.

comparative legal analysis, this article examines regulatory insufficiency in the face of digital violence. The results demonstrate that “moral panic” acts as a covert discursive mechanism, allowing hate speech to masquerade as legitimate concern to bypass corporate moderation and violate the dignity of historically marginalized groups, particularly women and LGTBIQ+ collectives. The article concludes that corporate self-regulation is unviable and demands the active intervention of the Mexican State. Based on the Inter-American principle that offline rights apply online, there is an urgent need for a regulatory framework that subordinates the algorithmic business model to the strict respect of human dignity.

Keywords: algorithms, attention economy, hate speech, moral panic, human rights.

1. Introducción

En el ecosistema digital contemporáneo, la configuración del discurso público ha sido desplazada hacia una infraestructura gobernada por la “economía de la atención”. En este escenario, los sistemas algorítmicos de las plataformas sociodigitales no operan bajo criterios de neutralidad axiológica, sino que optimizan el *engagement* mediante la priorización de contenidos polarizantes y extremos que logran retener al usuario en bucles de retroalimentación. Esta arquitectura del aislamiento, materializada en filtros burbuja y cámaras de eco, fragmenta el debate democrático y supedita la preservación de la dignidad humana a la rentabilidad estadística, creando nichos donde la otredad es percibida como una amenaza sistemática.

Bajo esta lógica de fragmentación social, emerge con vigor la fenomenología del “pánico moral”. El discurso alarmista opera como un mecanismo de infiltración discursiva que permite la proliferación de narrativas de odio bajo la apariencia de una preocupación legítima por la salvaguarda de los valores tradicionales o el bienestar colectivo. Al instrumentalizar el miedo y la indignación, estas expresiones logran eludir los mecanismos ordinarios de moderación de contenidos, institucionalizando una estigmatización que lesiona profundamente el honor y la reputación de grupos históricamente marginados.

Frente a este panorama, el presente artículo tiene como objetivo central analizar la insuficiencia del marco normativo mexicano ante los retos que impone la violencia digital, contrastándolo con los estándares internacionales de derechos humanos, con especial énfasis en el Plan de Acción de Rabat y la doctrina emanada de la Relatoría Especial para la Libertad de Expresión (RELE). Se busca evidenciar los vacíos legales que permiten la perpetuación de la discriminación algorítmica y la necesidad de una regulación que armonice la libertad de expresión con el principio de no discriminación.

Para alcanzar dicho propósito, la investigación se sustenta en una metodología de enfoque cualitativo con un alcance descriptivo-analítico. El estudio emplea el método estrictamente documental, tomando como andamiaje teórico la teoría de la acción comunicativa de Habermas (1987) para evaluar la racionalidad del discurso, así como las categorías de estigmatización propuestas por Waldron (2012). Complementariamente, se efectúa una revisión de literatura focalizada en las manifestaciones de violencia digital dirigida contra las mujeres y los colectivos LGBTQ+, articulada con el método jurídico-comparativo, indispensable para contrastar la legislación nacional vigente con el bloque de constitucionalidad interamericano y los tratados internacionales de los que México es parte.

En consecuencia, el trabajo se organiza en cinco ejes fundamentales: en primer lugar, se desentraña la mecánica algorítmica y su vínculo con la economía de la atención; en segundo término, se establecen los fundamentos teóricos del discurso y la dignidad; posteriormente, se aborda la fenomenología del pánico moral; se prosigue con un análisis documental centrado en las dinámicas de violencia digital hacia los colectivos señalados; y, finalmente, se presenta el estudio crítico del marco normativo y los desafíos regulatorios del Estado mexicano.

2. Mecánica algorítmica y economía de la atención: la arquitectura del aislamiento digital

2.1. Optimización del *engagement* y bucles de retroalimentación

Es de relevancia desentrañar el funcionamiento del algoritmo para concretar en la incidencia de discursos de odio a través de redes sociales. Un algoritmo es básicamente un sistema computacional que recopila información de los usuarios en plataformas digitales con el propósito de proveer un contenido personalizado que refleje el interés de consumo de las personas en el mundo digital (Metzler y García, 2024).

Entre las funciones principales de los algoritmos se encuentra el optimizar las estadísticas del *engagement*, que se conforma por la cantidad de interacciones con el contenido, ya sean *likes*, visualizaciones o comentarios. Este sistema busca retener a la persona en redes sociales por medio de bucles de retroalimentación; es decir, el algoritmo identifica la tendencia de interacción con un tipo de contenido, lo que posteriormente propicia que aparezcan materiales similares que capten la atención del consumidor (Metzler y García, 2024).

Si bien puede resultar bastante útil que las redes sociales se encuentren personalizadas en servicio de los usuarios, una realidad compleja es que los algoritmos ponderan contenido con potencial de masificarse sin importar las repercusiones sociales del mismo. El material que es posicionado con mayor frecuencia en las redes sociales guarda una connotación polarizante, controversial y extrema, esto debido a que provoca reacciones de sorpresa que incentivan la viralización de publicaciones (Suárez Poveda y Del Campo Saltos, 2024).

De acuerdo con Ging (2019), la superposición de ideologías como el neoliberalismo y el machismo hallan un nicho prolífico de la denominada lógica algorítmica por conducto de redes como YouTube, esto en atención a que los contenidos con connotaciones misóginas poseen un alto alcance en cuanto a su *engagement* ante las masas. Incluso, la creación de este contenido es recompensada económicamente con anuncios publicitarios de marcas que se exhiben en estos canales por su elevada visualización en la plataforma; entonces, estas publicaciones irreverentes constituyen una nueva forma de mercancía.

Las redes sociales y plataformas digitales suelen regirse por políticas que regulan la creación de espacios violentos o inapropiados, principalmente contra el llamado *hate speech*; no obstante, esta batalla entre la regulación de las plataformas y el *engagement* tiene como resultado que el control de materiales se vea desdibujado ante la priorización del contenido que sea especialmente rentable (Márquez y Ardèvol, 2018).

2.2. El “emburbujamiento” digital: filtro burbuja y cámaras de eco

Dentro de esta sofisticada estructura digital, existe un término denominado filtro burbuja que, de acuerdo con Pariser (2017) y López Julca *et al.* (2024), tiene lugar cuando la conducta humana es optimizada en las redes sociales con el fin de mejorar esta experiencia inmersiva. Este filtro genera que en redes sociales las personas solamente reciban contenido que tiene relación con sus intereses, creencias y gustos.

Este fenómeno resulta interesante en el sentido de que las personas no son conscientes de la magnitud con la que retroalimentan de datos a las plataformas digitales. Es decir, esta personalización del usuario no solamente se compone de aquello que se hace público en el internet, sino que los usuarios nutren este sistema de manera consciente e inconsciente, en donde, de una forma cada vez más intrusiva, las plataformas utilizan su información privada a manera de retenerlos con contenido de su preferencia.

Pariser (2017), uno de los mayores estudiosos del filtro burbuja, profundizó sobre los beneficios y retos que presenta esta personalización de los usuarios en redes sociales por medio de los algoritmos, donde destacó que, en una época en la que las personas se encuentran inmersas ante un abrumador tránsito de información en internet, la existencia de filtros personalizados resulta atractiva, ya que esto representa una opción cómoda que ofrece el contenido que se desea ver entre millones de opciones.

El actual funcionamiento de las *cookies* y los algoritmos ha logrado introducirse en la cotidianidad de la sociedad, permeando con facilidad y sin restricciones ante el desconocimiento o la falta de consentimiento de las personas de proporcionar su información personal. De acuerdo con Bajaña Tovar (2021), un aspecto complejo es que los principios de pluralidad, objetividad y neutralidad en el internet son desplazados para dar paso al “emburbujamiento” de los usuarios.

Los internautas, de una u otra forma, terminan siendo parte de la cámara de eco, figura que tiene lugar cuando personas con opiniones e intereses coincidentes son agrupados en segmentos en donde su discurso puede potencializarse y se mantienen al margen de las críticas. Este nicho digital se caracteriza por fomentar una sensación de “opinión masiva” o que “todas las personas opinan lo mismo”; esta supuesta hegemonía pudiera provocar la proliferación de discursos de odio (Rodríguez Cano, 2018).

En las cámaras de eco, como los individuos que las conforman tienen ideologías políticas homogéneas, se da un fenómeno que potencializa el discurso ya que solo escuchan su propia voz, como sucede con el efecto acústico del eco, en donde no hay alguien que opine diferente o cuestione los valores del grupo (Lombana-Bermudez *et al.*, 2022).

Otros estudios refieren a la intención del internet de construir burbujas de filtro a través de espacios en los que los internautas son agrupados en un sentido sociopolítico. Además, esta separación de grupos de personas apela directamente a los sentimientos de los usuarios al generarse un fuerte sentido de pertenencia a un grupo y una total fragmentación con otros (Waisbord, 2020).

Por esto último, existen posturas que asocian el fenómeno actual de la polarización a las cámaras de eco y las burbujas del filtro; precisamente, el internet forma parte del mercado, por tanto, no existe la intención de mostrar al internauta información nueva o desconocida que le permita incluso cuestionar su construcción de valores. Más bien, las redes sociales priorizan la permanencia de los usuarios alimentando sus “dietas de información” (Waisbord, 2020).

3. Teoría del discurso: racionalidad frente a la transgresión de la dignidad humana

3.1. El discurso racional y la acción comunicativa de Habermas

En este apartado será necesario partir de una conceptualización del discurso, para ello se recurre a la aportación de Habermas (1987), quien considera que la consolidación de un discurso racional amerita que la persona desde la vertiente cognitivo-intelectual emita opiniones fundadas, se encuentre dispuesta a aceptar los errores y esté en disposición a que su hipótesis sea cuestionada por otros sujetos.

La teoría de Habermas delimita además una serie de reglas para lograr una argumentación racional, en donde en la comunidad de diálogo todos deben concebirse con el mismo derecho de argumentar y prestarse a escuchar a los demás. Pese a que dicha tarea se vislumbra algo compleja, el autor considera que es más que un ideal, ya que oponerse a este valor democrático en la comunicación sería obstruir el proceso de la argumentación.

Con base en la teoría de la comunicación de Habermas, se entiende que el diálogo necesita de reciprocidad entre las partes, en donde los involucrados tengan la misma oportunidad de exponer sus puntos de vista de forma que sea proclive concretar una argumentación racional. Por tanto, cuando los planteamientos se emiten de manera unilateral, como sucede con las proclamas de odio, es posible considerar que se está ante un discurso aberrado al no cumplir con los requisitos de racionalidad en la comunicación.

3.2. El discurso de odio como difamación colectiva y ataque a la dignidad

Elaborar una definición conceptual del discurso de odio resulta una tarea ardua al no ser un concepto unívoco, por ende, lo mejor es identificar un elemento primordial dentro de estas expresiones. Por parte de Waldron (2012), es posible desentrañar el factor común en los discursos de odio: la capacidad que tiene el discurso de transgredir la dignidad humana de las personas.

La dignidad humana sería esa figura metafísica que facilita al individuo exigir a quienes le rodean el reconocimiento como sujeto apto para la vida en sociedad; es decir, es una reputación ante los ojos de los demás que le garantiza un trato equitativo en la comunidad (Díaz Soto, 2015).

Tomando como base a Waldron (2012), el discurso de odio podría traducirse como una difamación colectiva, situación en la que se asocian características de un colectivo específico con comportamientos antisociales. Para dicho autor, superando el aspecto iusnaturalista de la valía intrínseca al ser humano, aquella narrativa que atente contra la reputación de un sector social en menoscabo de su dignidad debe ser restringida con el objetivo de mantener una sociedad bien ordenada. A efecto de simplificar las modalidades en las que se dan estas expresiones, se expone la siguiente tabla:

Mecanismo Discursivo	Descripción Conceptual	Ejemplo Práctico
Imputación generalizada de ilícitos	Atribuir la comisión de delitos a la totalidad de un grupo poblacional sin evidencia individualizada.	Afirmar que todas las personas migrantes indocumentadas son responsables del aumento del crimen organizado o el narcotráfico.
Caracterización denigrante	Asignar atributos morales o intelectuales deficientes basados en prejuicios históricos.	Estigmatizar a una minoría étnica o religiosa afirmando que son inherentemente deshonestos, fanáticos o inferiores intelectualmente.
Deshumanización	Despojar a una colectividad de su dignidad humana mediante el uso de lenguaje degradante.	Referirse a un grupo social vulnerable como una “plaga”, una “infección” o compararlos con parásitos.
Segregación explícita	Promover la negación de derechos fundamentales o el bloqueo a espacios físicos e institucionales.	Impulsar campañas para negar el acceso a la atención médica, la vivienda o la educación a individuos por su origen nacional.

Tabla 1. *Mecanismos de estigmatización en el discurso de odio.* Fuente: Elaboración del autor con información proporcionada por Waldron (2012).

Las aportaciones de Waldron respecto a los mecanismos en los que se construyen los discursos de odio permiten remontarse en el pasado, con acontecimientos trágicos como el Holocausto Nazi o el Genocidio de Ruanda. En este tipo de acontecimientos se observa que la polarización de la otredad puede propiciar las vulneraciones sistemáticas a los derechos humanos de un grupo determinado de acuerdo con supuestas características.

En esta intención de vivir en una sociedad bien ordenada, es fundamental que, pese a las diferencias, las reglas jurídicas estén al servicio de la protección de la dignidad humana; la violación a este elemento primordial en los derechos fundamentales de un colectivo específico presupone vulneraciones futuras a otros sectores, lo que terminaría por corromper el Estado de derecho y el orden social. La dignidad humana constituye una figura que solamente puede funcionar plenamente bajo la universalidad. Ahora bien, la Organización de las Naciones Unidas (ONU) ofrece su propio concepto del discurso de odio:

Cualquier forma de comunicación de palabra, por escrito o a través del comportamiento, que sea un ataque o utilice lenguaje peyorativo o discriminatorio en relación con una persona o un grupo sobre la base de quiénes son o, en otras palabras, en razón de su religión, origen étnico, nacionalidad, raza, color, ascendencia, género u otro factor de identidad. (ONU, 2019, p. 3)

Partiendo de esta cita, es posible concretar correctamente que la denostación a la identidad de las personas es utilizada con una connotación negativa y discriminatoria, justificando este lenguaje ofensivo con un ápice de “justicia”. O sea, los discursos de odio sostienen una supuesta verdad contra algún grupo en particular en degradación de su reputación y dignidad humana.

4. Fenomenología del pánico moral: el discurso alarmista como mecanismo de infiltración

4.1. La demonización de la otredad y las guerras culturales

Cuando en una sociedad determinada comienzan a quebrantarse las estructuras tradicionales o hegemónicas, se labra el discurso alarmista, que da paso a la preocupación o al escándalo ante el temor de abandonar el *statu quo* de una comunidad. Según Pecourt, Rubio Arostegui y Aranda (2024), cuando se obstruyen los procesos tradicionales de incorporación, existe un choque entre dos polos ideológicos por medio de las guerras culturales, mismas que traen consigo la demonización de ciertos grupos poblacionales.

El pánico moral constituye un fenómeno sociológico caracterizado por una reacción masiva y desproporcionada de la sociedad frente a una condición, episodio o grupo de personas que son definidos como una amenaza para los valores e intereses sociales. Como ha sido conceptualizado clásicamente por Cohen ([1972] 2011) en su estudio fundacional, esta fenomenología requiere la construcción de un “demonio popular” (*folk devil*), al cual se le señala como causante de la degradación moral y hacia el cual se canaliza la hostilidad colectiva.¹

Teóricamente, el pánico moral opera mediante la simplificación de problemas complejos y la creación de narrativas binarias que distinguen radicalmente entre un “nosotros” normativo y un “ellos” desviado. Con la información analizada, es posible entender la sinergia de factores que confluyen en la creación de los discursos de odio. Es decir, las diferencias culturales e ideológicas fomentadas por la exposición en los medios de comunicación –y actualmente amplificadas por las plataformas sociodigitales– propician la construcción sistemática de discursos de odio en contra de grupos no hegemónicos que no se adecúan a la estructura social tradicional.

4.2. Infiltración discursiva: eludiendo la moderación algorítmica

Resulta evidente que el pánico moral transita y se retroalimenta en torno a las lógicas algorítmicas; al profundizar en este fenómeno, se dilucida que los discursos alarmistas suelen obtener métricas muy elevadas en cuanto a *engagement*. Si bien los discursos de odio y alarmistas han existido históricamente, hoy han logrado colocarse en el centro de la interacción sociodigital debido a que la arquitectura de cámaras de eco y filtros burbuja constituye el espacio idóneo para la proliferación de expresiones polarizantes.

En este entorno, los discursos alarmistas funcionan como un mecanismo de infiltración o táctica de enmascaramiento frente a las restricciones normativas de internet. Logran posicionarse de manera sutil en las tendencias de las plataformas al disfrazarse de preocupación legítima por el bienestar común, lo que les permite eludir las políticas de moderación corporativa. De forma paulatina y discreta, estas narrativas construyen un andamiaje para el discurso de odio que sortea la censura automatizada al no presentar, en apariencia, un vocabulario explícitamente infractor.

5. Estudio de casos específicos: resistencia ante los derechos de las mujeres y colectivos LGTBIQ+

5.1. La misoginia digital y la mercantilización del odio sexista

Como se ha argumentado, frente a las estrategias alarmistas que operan como prácticas discursivas de odio bajo la justificación de un pánico moral, existe una tendencia sistemática a vulnerar a ciertos grupos históricamente marginados. Para efectos del presente análisis documental, se tomará como foco a los colectivos de mujeres y LGTBIQ+.

Estos grupos han adquirido mayor visibilidad tras la consolidación de ciertos derechos fundamentales. Resulta imperativo considerar, en retrospectiva, que los integrantes de estos colectivos sufrieron exclusión, discriminación y violencia de forma institucionalizada, motivo por el cual los Estados, bajo una agenda de derechos humanos, están obligados a brindarles protección prioritaria. La problemática contemporánea surge cuando, ante el avance de estos derechos, emerge una reacción conservadora de ciertos sectores sociales frente al cambio. En este contexto, las redes sociales se configuran como un escenario desregulado donde estos discursos de hostilidad pueden desplegarse y fortalecerse con un alto grado de impunidad.

Para Tortajada y Vera (2021), el internet se erige como un espacio hostil para las mujeres, ya que constantemente se documentan actos de intimidación, amenazas, acoso, insultos y extorsión. El entorno digital funciona como un catalizador de sesgos y prejuicios que, si bien en la interacción física pueden estar inhibidos por la convención social, en internet propician la reproducción de violencia histórica amparada, en gran medida, por el anonimato.

El contenido sexista y misógino es recurrente en plataformas de alojamiento de videos como YouTube, donde, bajo la apariencia de humor o sarcasmo, ciertos creadores de contenido alcanzan la viralidad frente a audiencias predominantemente jóvenes. Este fenómeno resulta sumamente rentable para el modelo de negocio digital; por ende, este tipo de discurso, amplificado por el algoritmo, se convierte en un producto económicamente conveniente (Tortajada y Vera, 2021).

5.2. La construcción de la “ideología de género” como dispositivo de pánico moral

Para Rossel-Castagneto (2025), los entornos digitales contemporáneos representan una amenaza para la dignidad humana, la igualdad y el pluralismo democrático, principalmente frente a los discursos de odio dirigidos contra poblaciones históricamente vulneradas, tales como personas migrantes, miembros de la comunidad LGTBIQ+ y mujeres.

El avance del feminismo y de las demandas de las disidencias sexuales puede detonar reacciones articuladas mediante tácticas de enmascaramiento discursivo. Al clasificar a estos movimientos sociales como amenazas directas a la familia, las buenas costumbres o el desarrollo infantil, se construyen narrativas alarmistas que ganan tracción gracias a las cámaras de eco y la amplificación algorítmica. Por ello, resulta indispensable evaluar los retrocesos materiales que el discurso de odio sociodigital puede infligir sobre la dignidad y los derechos humanos.

El debate en torno a la categoría peyorativa de “ideología de género” antecede a la era algorítmica. Este término fue gestado e impulsado inicialmente por la jerarquía eclesíastica católica y los movimientos neoconservadores como un dispositivo crítico frente a las posturas que cuestionaban la heteronormatividad y el esencialismo biológico, pilares que dichas visiones postulan como base exclusiva de la estructura familiar (Bárceñas Barajas, 2021).

La pertinencia de analizar la hostilidad hacia los colectivos LGTBIQ+ y los movimientos feministas radica en su instrumentalización dentro del debate político contemporáneo. Estos grupos son frecuentemente convertidos en el epicentro de la polarización política e ideológica, evidenciando una tensión constante entre la ampliación y el retroceso de los derechos humanos. Esta pugna se recrudece severamente bajo infraestructuras algorítmicas que fomentan el sesgo de confirmación y recompensan la confrontación discursiva.

Históricamente, desde la Cuarta Conferencia Mundial sobre la Mujer (Pekín, 1995) impulsada por Naciones Unidas, existió una profunda incomodidad por parte de la alta jerarquía católica, la cual alertó sobre una presunta imposición de agendas feministas. Ya para el año 2004, mediante documentos oficiales, dicha institución formalizó su preocupación ante los movimientos que refutaban la jerarquía “natural” entre hombres y mujeres, argumentando que esto comprometería los valores fundacionales de la familia y el orden social (Cornejo-Valle y Pichardo, 2017). Es en este último argumento donde el pánico moral opera como un encubrimiento retórico: se disfraza la invalidación de derechos como una defensa necesaria ante una supuesta amenaza existencial.

Bárceñas Barajas (2021) documenta, mediante un análisis en la plataforma Facebook sobre el Frente Nacional de la Familia (FNF), cómo este grupo conceptualiza la mal llamada “ideología de género” como una imposición dictatorial, una perversión sexual y un ataque a la naturaleza. Para los usuarios inmersos en este ecosistema digital, se instaura la percepción de estar librando una defensa moral legítima.

En consecuencia, la convergencia entre un entorno digital con deficiente regulación normativa y sistemas algorítmicos orientados a priorizar contenidos incendiarios resulta en la exacerbación y perpetuación del ciberacoso y el odio sistemático. Por tanto, es apremiante que los Estados asuman una agenda robusta de protección de los derechos humanos en el ciberespacio, garantizando especialmente el derecho a la igualdad y a la no discriminación.

Es posible concluir que el neoconservadurismo en México ha instrumentalizado el pánico moral para proyectar a las mujeres que ejercen su autonomía reproductiva y a las personas LGTBIQ+ como peligros inminentes para el tejido social. Esto se consolida mediante la fabricación de un escenario de posverdad, sostenido en *fake news* y campañas de desinformación que socavan el imaginario colectivo y anclan estigmas profundos sobre estos sectores poblacionales (Bárceñas Barajas, 2021).

6. Análisis normativo y estándares internacionales: retos de la regulación en la era digital

6.1. Estándares Universales e Interamericanos: del Plan de Rabat a la doctrina de la RELE

Con lo que se desarrolló durante el trabajo queda explícito el peligro que representa para la materialización de los derechos humanos la violencia digital que se ve fortalecida de la estructura algorítmica y la ausencia de regulación de contenido. Entonces, es fundamental analizar el cuerpo normativo que existe respecto a este fenómeno, o bien, encontrar los vacíos legales que presentan un área de oportunidad ante dicha problemática.

Dentro del Sistema Universal de los Derechos Humanos, por conducto del Pacto Internacional de Derechos Civiles y Políticos (PIDCP) en su artículo 20, le finca la responsabilidad a los Estados de los cuerpos normativos en que deben prohibir la incitación al odio nacional, religioso o racial que incite a la discriminación y la violencia (ONU, 1966).

Los derechos civiles y políticos requieren para su goce que no exista una segregación de ciertos grupos de la sociedad, ya que los mismos no se pueden restringir por características particulares. Con una redacción similar (que evidencia cierta inspiración con en el PIDCP), el Pacto de San José en su artículo 13.5 dispone la prohibición en los Estados parte de la apología del odio por cualquier motivo, a lo que, al estar suscrito, el Estado mexicano debe dar cumplimiento (Organización de los Estados Americanos, 1969).

En 2012, en una resolución del Consejo de Derechos Humanos de la ONU, se explicó la doble dimensión de los derechos humanos, con una interpretación que considera que estas potestades no solamente deben ser protegidas tanto en el ámbito físico, como en el digital; en este sentido, bajo la línea de este trabajo, la discriminación por medio del internet debería implicar una obligación del Estado.

Uno de los avances más destacados en el rubro de los derechos humanos en la vertiente digital se dio con el Plan de Acción de Rabat en la ONU (Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos [ACNUDH], 2012), mismo que, cabe precisar, es un documento orientador que constituye una especie de *soft law* y no una norma vinculante para los Estados.

En este documento se precisó el fenómeno de la multiculturalidad y la diversidad de medios de comunicación con la llegada de las redes sociales, mismo que ha permitido a la sociedad estar interconectada, además de una amplia difusión de información, lo que presupone un reto para los derechos humanos (ACNUDH, 2012).

El aspecto principal del plan es que insta a los Estados a la instauración de un marco regulador que permita la pluralidad y diversidad de medios de comunicación. Este principio es contrario a lo que sucede con el algoritmo que prioriza el *engagement* y crea cámaras de eco en donde las personas no son cuestionadas, al contrario, solamente reciben retroalimentación en el mismo sentido (ACNUDH, 2012).

La organización consideró que las plataformas digitales deben actuar de buena fe y autorregularse; no obstante, con la información revisada anteriormente, es posible observar que estos espacios virtuales circulan dentro del mercado, por lo que la permanencia de las personas en sus interfaces supone una ganancia económica (ACNUDH, 2012).

Entonces, en ese ámbito parece difícil lograr que los medios digitales sean más estrictos con la emisión de contenidos de odio. Además, se recalcó en que esta lucha contra la discriminación no basta con la regulación normativa e incluso las sanciones penales, sino que amerita el apoyo de la sociedad y de las plataformas digitales (ACNUDH, 2012).

En el ámbito interamericano, la Relatoría Especial para la Libertad de Expresión (RELE) de la Comisión Interamericana de Derechos Humanos (CIDH), emitió un informe en 2013 donde retoma el principio que ya venía desarrollándose en el espectro universal con la máxima “lo *offline* aplica *online*”, o sea, que la protección de los derechos humanos también tiene alcance en el mundo digital (RELE, 2013).

En su informe de 2017, la RELE fue aún más allá al señalar que los intermediarios tecnológicos poseen la obligación de adecuar sus algoritmos conforme a los principios rectores de los derechos humanos, en tanto controlan el monopolio de la moderación del debate público. Priorizar estas manifestaciones intolerantes debido a su rentabilidad económica contraviene el artículo 13.5 de la Convención Americana, que prohíbe la apología del odio (RELE, 2017).

6.2. El Ordenamiento Mexicano: insuficiencias y límites de la libertad de expresión

Con una mayor especificidad, en México se posee la Ley Federal para Prevenir y Erradicar la Discriminación (2003), misma que otorga conceptos respecto a este fenómeno como lo es la misoginia, la transfobia y la homofobia. Sin embargo, aún no se ha reformado de tal manera que englobe la discriminación ejercida a través de redes sociales y el factor del algoritmo que perpetúa los discursos de odio en el mundo digital.

La paradigmática reforma constitucional del año 2011 en materia de derechos humanos estableció el derecho a la no discriminación en el artículo 1 constitucional, esto como una forma de proteger la dignidad humana como pilar primigenio de los derechos fundamentales (Constitución Política de los Estados Unidos Mexicanos, 1917).

La Constitución, como norma suprema en conjunto con el bloque de constitucionalidad, debe constituir siempre en México un escudo para todos los grupos en situación de vulnerabilidad. En este rubro, resulta imprescindible destacar que el Estado mexicano ha ratificado instrumentos clave del Sistema Interamericano que robustecen estas garantías, específicamente la Convención Interamericana contra el Racismo, la Discriminación Racial y Formas Conexas de Intolerancia (A-68) y la Convención Interamericana contra toda Forma de Discriminación e Intolerancia (A-69).

En ambas convenciones, el artículo 4, fracción II, relativo a los “Deberes de los Estados”, establece el compromiso ineludible de los Estados Parte de prevenir, eliminar, prohibir y sancionar toda publicación, circulación o diseminación de materiales, en cualquier medio de comunicación, que promuevan o inciten al odio, la discriminación y la intolerancia. En consecuencia, esta cláusula pétrea constitucional y su ampliación convencional no permiten que se legitime ningún tipo de segmentación ni estigmatización dentro del Estado, haciendo obligatoria la transposición de estas obligaciones protectoras al ciberespacio y a las plataformas digitales.

En México, el derecho a la libertad de expresión está claramente delimitado en el artículo 6 constitucional; sin embargo, existen limitaciones en el ejercicio de este: la expresión de ciertas ideas no debe vulnerar derechos de terceros, atentar contra la moral o constituir algún delito (Constitución Política de los Estados Unidos Mexicanos, 1917). Por ello, los discursos de odio que buscan restringir derechos o adjudicar características negativas a cierto grupo a manera de preocupación legítima constituyen una violación a los límites de la libertad de expresión y a los compromisos internacionales asumidos por el país.

7. Conclusiones

El análisis documental desarrollado a lo largo de esta investigación permite concluir, en primer término, que delegar la moderación del discurso exclusivamente a la “autorregulación” corporativa presenta severas limitaciones operativas. Si bien las plataformas sociodigitales pueden implementar políticas internas y actuar de buena fe en ciertos contextos, no son intermediarios neutros; su modelo de negocio se encuentra estructuralmente ligado a la retención del usuario. Por ende, la autorregulación resulta insuficiente por sí sola frente a una “economía de la atención” donde el contenido extremista o polarizante suele generar altos niveles de interacción y, en consecuencia, una mayor rentabilidad.

En este sentido, la evidencia sugiere que la arquitectura algorítmica sostenida en las “cámaras de eco” y los “filtros burbuja” tiende a aislar a los usuarios, fragmentando el debate democrático y limitando el pluralismo. Al dificultar el encuentro con la otredad y, a menudo, priorizar la confrontación, estas infraestructuras digitales obstaculizan el discurso racional y abren espacios para que el pánico moral opere como un mecanismo encubierto que perpetúa violencias simbólicas y materiales contra grupos históricamente marginados, como las mujeres y los colectivos LGTBIQ+.

Frente a esta asimetría de poder, y con base en el marco normativo analizado, el Estado mexicano requiere asumir un rol más proactivo y garante. El vacío regulatorio actual propicia un escenario en el que la moderación de la esfera pública puede quedar supeditada a intereses mercantiles, lo cual entra en tensión con las obligaciones constitucionales e internacionales del país. Conforme a los estándares del Sistema Interamericano y la doctrina de la RELE, de conformidad con el hecho de que los derechos que las personas tienen en el entorno físico deben ser igualmente protegidos en el entorno digital, es pertinente la intervención estatal. La libertad de expresión, delineada en el artículo 6 constitucional, no está diseñada para amparar la masificación algorítmica de la difamación colectiva.

Por consiguiente, es recomendable transitar hacia un paradigma legislativo e institucional más robusto y complementario. La evidencia documental apunta a que el Estado de derecho debe aportar lineamientos normativos claros que exijan mayor transparencia y auditoría en los diseños algorítmicos, complementando así la moderación corporativa. La exigencia contemporánea indica que resulta necesaria la creación de un marco normativo que armonice la innovación y la economía de la atención con la protección transversal de la dignidad humana, garantizando que el entorno digital opere bajo un respeto irrestricto a los derechos fundamentales.

Referencias bibliográficas

- Bajaña Tovar, Fernando. (2021). Filtro burbuja: ¿Cuál es el costo de la personalización digital? *Revista Chilena de Derecho y Tecnología*, 10(1), 29-52. <https://doi.org/10.5354/0719-2584.2021.54042>
- Bárceñas Barajas, Karina. (2021). La violencia simbólica en el discurso sobre la 'ideología de género': una perspectiva desde la dominación simbólica a través del pánico moral y la posverdad. *Intersticios Sociales*, (21), 125-150. <https://tinyurl.com/2cjybpqd>
- Cohen, Stanley. (2011 [1972]). *Folk devils and moral panics: the creation of mods and rockers*. Routledge.
- Constitución Política de los Estados Unidos Mexicanos. (1917, 5 de febrero). *Constitución Política de los Estados Unidos Mexicanos*. Diario Oficial de la Federación.
- Cornejo-Valle, Mónica y Pichardo, J. Ignacio. (2017). La "ideología de género" frente a los derechos sexuales y reproductivos. El escenario español. *Cadernos Pagu*, (50), e175009. <https://doi.org/10.1590/18094449201700500009>
- Díaz Soto, José Manuel. (2015). Una aproximación al concepto de discurso del odio. *Revista Derecho del Estado*, (34), 77-101. <https://doi.org/10.18601/01229893.n34.05>
- Ging, Debbie. (2019). Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and Masculinities*, 22(4), 638-657. <https://doi.org/10.1177/1097184X17706401>
- Habermas, Jürgen. (1987). *Teoría de la acción comunicativa*. Vol. I. Taurus.
- Ley Federal para Prevenir y Erradicar la Discriminación. (2003, 11 de junio). *Ley Federal para Prevenir y Erradicar la Discriminación*. Diario Oficial de la Federación.
- Lombana-Bermúdez, Andrés; Vallejo Mejía, Maryluz; Gómez Céspedes, Liliana María y Pino Uribe, Juan Federico. (2022). Cámaras de eco, desinformación y campañas de desprestigio en Colombia. Un estudio de Twitter y las elecciones locales de Medellín en 2019. *Política y Gobierno*, 29(1), 1-30. <https://tinyurl.com/224er9tm>
- López Julca, Carlos; Julca Guerrero, Félix; Nivin Vargas, Laura; Allauca Castillo, Wendy; Robles Trejo, Luis y Robles Blacido, Elmer. (2024). El filtro burbuja y el derecho a la información en la web. *Desde el Sur*, 16(1), 1-21. <https://tinyurl.com/23w97wss>
- Márquez, Israel y Ardèvol, Elisenda. (2018). Hegemonía y contrahegemonía en el fenómeno youtuber. *Desacatos*, (56), 34-49. <https://tinyurl.com/2dz5hhne>
- Metzler, Hannah y García, David. (2024). Social Drivers and Algorithmic Mechanisms on Digital Media. *Perspectives on Psychological Science*, 19(5), 735-748. <https://doi.org/10.1177/17456916231185057>
- Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos. (2012, 5 de octubre). Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. En *Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred* (pp. 6-15). Organización de las Naciones Unidas. [ACNUDH] <https://tinyurl.com/26q2xahe>
- Organización de las Naciones Unidas. (1966, 16 de diciembre). *Pacto Internacional de Derechos Civiles y Políticos*. Organización de las Naciones Unidas. [ONU]

- Organización de las Naciones Unidas. (2019). *Estrategia y plan de acción de las Naciones Unidas para la lucha contra el discurso de odio*. Organización de las Naciones Unidas. [ONU] <https://tinyurl.com/24u7a35g>
- Organización de los Estados Americanos. (1969, 22 de noviembre). *Convención Americana sobre Derechos Humanos "Pacto de San José de Costa Rica"*. Organización de los Estados Americanos.
- Pariser, Eli. (2017). *El filtro burbuja: Cómo la red decide lo que leemos y lo que pensamos*. Taurus.
- Pecourt, Juan; Rubio Arostegui, Juan Arturo y Aranda, José María. (2024). Las guerras murales. Feminismo popular y pánico moral en Madrid: un estudio de caso del barrio de La Concepción. *RIPS: Revista de Investigaciones Políticas y Sociológicas*, 23(2), 42-62. <https://doi.org/10.15304/rips.24.2.9876>
- Relatoría Especial para la Libertad de Expresión. (2013). *Libertad de expresión e Internet*. Comisión Interamericana de Derechos Humanos. [RELE]
- Relatoría Especial para la Libertad de Expresión. (2017). *Estándares para una Internet libre, abierta e incluyente*. Comisión Interamericana de Derechos Humanos. [RELE]
- Rodríguez Cano, Carlos Antonio. (2018). Los usuarios en su laberinto: burbujas de filtros, cámaras de eco y mediación algorítmica en la opinión pública en línea. *Virtualis*, 8(16), 57-76. <https://tinyurl.com/239eogjm>
- Rossell-Castagneto, María Lorena. (2025). Regular el odio digital: entre la libertad de expresión y la protección de colectivos vulnerables en Chile. *Universitas XXI*, 43, 95-122. <https://doi.org/10.17163/uni.n43.2025.04>
- Suárez Poveda, José Francisco y Del Campo Saltos, Guillermo. (2024). Algoritmos en redes sociales y estrategias comunicacionales para el posicionamiento de medios tradicionales. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 5(6), 1670-1686. <https://doi.org/10.56712/latam.v5i6.3114>
- Tortajada, Iolanda y Vera, Teresa. (2021). Presentación del monográfico: Feminismo, misoginia y redes sociales. *Investigaciones Feministas*, 12(1), 1-4. <https://doi.org/10.5209/infe.74446>
- Waisbord, Silvio. (2020). ¿Es válido atribuir la polarización política a la comunicación digital? Sobre burbujas, plataformas y polarización afectiva. *Revista SAAP*, 14(2), 249-279. <https://doi.org/10.46468/rsaap.14.2.a1>
- Waldron, Jeremy. (2012). *The harm in hate speech*. Harvard University Press.

NOTAS

- ¹ Cabe precisar que la referencia bibliográfica de Cohen remite a la obra original de 1972, momento en que acuñó el término, si bien para el presente análisis se ha consultado la edición revisada de 2011, la cual confirma la vigencia de estos mecanismos de estigmatización.

AmeliCA

Disponible en:

<https://portal.amelica.org/ameli/ameli/journal/785/7855782009/7855782009.pdf>

Cómo citar el artículo

Número completo

Más información del artículo

Página de la revista en portal.amelica.org

AmeliCA

Ciencia Abierta para el Bien Común

Cristian Alberto López Rodríguez

**PÁNICO MORAL, DISCURSOS DE ODIO Y EL ASEDIO
ALGORÍTMICO A LOS DERECHOS HUMANOS**

Hate Speech, and the Algorithmic Siege on Human Rights

Cuadernos de Literatura

núm. 30, e3007, 2026

Universidad Nacional del Nordeste, Argentina

revistas@unne.edu.ar

ISSN: 0326-5102

ISSN-E: 2684-0499

DOI: <https://doi.org/10.30972/clt.309673>