



**Algoritmos y Silenciamiento Digital:**

Invisibilización de los discursos minoritarios en la era tecnológica

**Autoras**

Maria Fernanda Cuevas Rivera

Manuella Velásquez Ramirez

**Director**

Julio Cesar Gaitán Bohorquez

**Abogadas**

**Facultad de Jurisprudencia**

**Programa de Jurisprudencia.**

**Semillero de investigación: Derechos Humanos en Entornos Digitales**

**Universidad del Rosario**

**Bogotá - Colombia**

**2026**

# ÍNDICE

<b>I. RESUMEN.....</b>	<b>3</b>
<b>II. ABSTRACT.....</b>	<b>4</b>
<b>III. INTRODUCCIÓN.....</b>	<b>5</b>
3.1 Definición del Problema.....	6
3.2 Propósito de investigación e hipótesis.....	7
3.3 Marco teórico.....	8
<b>IV. METODOLOGÍA.....</b>	<b>11</b>
4.1 Diseño del estudio:.....	11
4.2 Método de análisis:.....	11
4.3 Limitaciones metodológicas encontradas:.....	12
<b>V. ARGUMENTACIÓN.....</b>	<b>12</b>
5.1 El algoritmo.....	12
5.2. Neutralidad algorítmica inexistente.....	14
5.3. El algoritmo y la libertad de expresión.....	17
<b>VI. CONCLUSIÓN.....</b>	<b>30</b>
6.1 Actualidad.....	30
6.2 Iniciativas en el marco regulatorio y normativo.....	31
6.3 La necesidad.....	32
<b>VII. BIBLIOGRAFÍA.....</b>	<b>33</b>

## I. RESUMEN

El artículo aquí presentado, examinará y analizará el fenómeno relacionado entre el silenciamiento digital y algorítmico, propiciado por las plataformas tecnológicas masivas, y la sistemática invisibilización de discursos que provienen de comunidades que se han considerado históricamente marginadas y minoritarias. El objetivo es producir un punto de vista alterno, fundamentado en un juicioso análisis cualitativo, de cómo el contenido digital tiene un fuerte impacto en la represión y restricción del ejercicio de la libertad de expresión y la libertad, cuyo origen parte de la base de la moderación algorítmica automatizada y controlada.

La metodología utilizada fue la revisión de fuentes cualitativas, que a través de la revisión de documentos, estudio de jurisprudencia y casos empíricos, normativa internacional en materia de digitalización y derechos humanos, junto la revisión con informes de diferentes instituciones a nivel regional, nacional e internacional, visibilizan los casos en que los algoritmos ponen en tela de juicio a las minorías, teniendo una dinámica discriminatoria y represiva. Con ello, este artículo resalta entonces esa notoria tensión entre el mecanismo de control de masas, tal y como funciona a través de la digitalización de algoritmos, y las obligaciones de los poderes de los estados en proteger, humanizar y dignificar a todas las comunidades dentro de un modelo social.

Gracias a este minucioso estudio, será posible confirmar la creación de un sesgo generacional creado a raíz de la estructura tecnológica que se implementa en la actualidad. Con ello, las grandes industrias tecnológicas son encargadas de introducirlos a la sociedad, produciendo sistemas algorítmicos que insertan en las comunidades minoritarias, una posición aún más desventajosa. Así, estos modelos propenden a suprimir cierto contenido, ocultarlo, desviarlo, haciendo que su visualización se torne escasa o, por el contrario, que su amplificación se dirija a un público que no muestra interés en este tipo de aspectos.

El presente artículo no solo concluye con la evidencia de censura al discurso minoritario debido a la configuración del algoritmo, sino que también propone iniciativas regulatorias que deben incorporarse en las normativas vigentes. Este fenómeno, e incluyendo en general la creciente digitalización y rápida creación de nuevas tecnologías, debe encontrar su fuente en un sistema legislativo riguroso, dotado de transparencia algorítmica, permeado de responsabilidad y uso responsable de la información pero, sobre todo, que cuente con una inyección, no solo teórica sino también material, de estándares de derechos humanos y gobernanza algorítmica que sea capaz de tutelar de la manera más eficiente posible, la diversidad de voces en el entorno tecnológico.

**Palabras clave:** algoritmos, silenciamiento digital, sesgo algorítmico, libertad de expresión, discursos minoritarios, moderación de contenidos, *shadow banning*, gobernanza algorítmica, derechos humanos digitales, burbujas de filtro.

## II. ABSTRACT

The article herein presented will examine and analyse the interrelated phenomenon of digital and algorithmic silencing—fostered by massive technology platforms—and the systematic invisibilisation of discourses originating from communities that have been historically considered marginalised and minoritised. The objective is to produce an alternative perspective, grounded in a rigorous qualitative analysis, on how digital content exerts a significant impact on the repression and restriction of the exercise of freedom of expression and liberty, stemming from the foundations of automated and controlled algorithmic moderation.

The methodology employed in this article is qualitative in nature, drawing upon document review, case law analysis and empirical cases, international regulations concerning digitalisation and human rights, together with reports from various institutions at the regional, national, and international levels, in order to bring to light instances in which algorithms call into question the standing of minorities through discriminatory and repressive dynamics. In doing so, it underscores the manifest tension between the mechanism of mass control—as it operates through the digitalisation of algorithms—and the obligations incumbent upon state powers to protect, humanise, and dignify all communities within a social model.

Through this meticulous study, it will be possible to confirm the emergence of a generational bias created as a consequence of the technological infrastructure currently being implemented. In this regard, major technology corporations are responsible for introducing such systems into society, producing algorithmic architectures that place minoritised communities in an even more disadvantageous position. Accordingly, these models tend to suppress certain content, conceal it, or divert it, rendering its visibility scarce or, conversely, directing its amplification towards audiences that display no interest in such matters.

This article concludes by proposing—and indeed seeking—regulatory initiatives that must begin to be incorporated into existing legal frameworks. This phenomenon, together with the broader trend of increasing digitalisation and the rapid development of new technologies, must find its foundation in a rigorous legislative system endowed with algorithmic transparency, permeated by accountability and responsible use of information, and above all, infused—not only theoretically but also materially—with human rights standards and algorithmic governance capable of safeguarding, in the most effective manner possible, the diversity of voices within the technological environment.

**Keywords:** algorithms, digital silencing, algorithmic bias, freedom of expression, minority discourses, content moderation, *shadow banning*, algorithmic governance, digital human rights, filter bubbles.<sup>1</sup>

---

<sup>1</sup> La traducción de este abstract fue construido con ayuda de Inteligencia Artificial, haciendo uso del siguiente prompt: “hazme la traducción oficial de esto sin perder el sentido y traduciendo de forma perfecta: El artículo aquí presentado, examinará (...)”

### III. INTRODUCCIÓN

Actualmente, las plataformas digitales son escenarios de comunicación, entretenimiento, consumo y debate. Sin embargo, más allá de estas funciones, tal y como señalan Van Dijck et al. (2018), hoy en día, se vive en una “sociedad de plataformas” en la que las infraestructuras tecnológicas no se limitan a medir interacciones (v.g. likes, followers, etc), sino que determinan quién puede hablar, y de forma más importante, quién puede ser escuchado. En este sentido, redes sociales como Tik Tok, Instagram, X (antiguamente twitter), Youtube, entre otras; dominan el tiempo de conexión y controlan gran parte de la información a nivel mundial.

En este contexto, los algoritmos desempeñan un papel trascendental en la estructuración del mundo digital, entendido como "aquello que abarca lo relacionado con los ordenadores, Internet o las redes de comunicación" (Telefónica, 2025). De acuerdo con el profesor Alejandro Rodríguez de la Universidad Politécnica de Madrid, son estos quienes definen la información de las personas para optimizar la experiencia del usuario. Sin embargo, estos no son neutrales, ya que pueden “generar una visión distorsionada de la realidad y la percepción de que las opiniones propias son mayoritarias o incluso las únicas” (Universidad Politécnica De Madrid, 2025). En consecuencia, los algoritmos, no se limitan a la organización de la información en el ecosistema digital, pues lo que ocurre en este entorno tiene efectos directos sobre la vida social fuera de las pantallas; de ahí que también influyan en la construcción de prejuicios sociales.

Adicionalmente, las infraestructuras digitales son espacios que facilitan el ejercicio del derecho fundamental de la libertad de expresión, ya que permiten a los individuos a obtener, producir y divulgar información de forma masiva (Calcáneo, 2021). Empero, esta “democratización” del discurso se encuentra suscitada por lógicas algorítmicas que pueden llegar a favorecer ciertos contenidos sobre otros (Jaramillo, s. f. ), lo que puede conllevar a dinámicas de silenciamiento digital o visibilidad diferenciada.

Ahora bien, desde una visión jurídica, este fenómeno pone en discusión de forma directa el análisis y alcance del derecho a la libertad de expresión. Este se encuentra resguardado en diferentes cuerpos normativos: en primer lugar, a nivel internacional el artículo 13 de la Convención Interamericana sobre Derechos Humanos (CIDH, 1969) y el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos, respectivamente:

*“1. Toda persona tiene derecho a la libertad de pensamiento y de expresión. Este derecho comprende la libertad de buscar, recibir y difundir informaciones e ideas de toda índole, sin consideración de fronteras, ya sea oralmente, por escrito o en forma impresa o artística, o por cualquier otro procedimiento de su elección.” (CIDH, 2000, art 13)*

*“(...) Toda persona tiene derecho a la libertad de expresión; este derecho comprende la libertad de buscar, recibir y difundir informaciones e ideas de toda índole, sin consideración de fronteras, ya sea oralmente, por escrito o en forma impresa o artística, o por cualquier otro procedimiento de su elección. (...)” (ONU, 1966, art 19)*

En segundo lugar, a nivel nacional en el artículo 20 de la Constitución Política de Colombia de 1991:

*“ARTÍCULO 20. Se garantiza a toda persona la libertad de expresar y difundir su pensamiento y opiniones, la de informar y recibir información veraz e imparcial, y la de fundar medios masivos de comunicación. Estos son libres y tienen responsabilidad social. Se garantiza el derecho a la rectificación en condiciones de equidad. No habrá censura.” (Constitución Política de Colombia, 1991, art. 20)*

A pesar de contar con una “protección” normativa, la mediación algorítmica<sup>2</sup> incluye nuevos modelos de restricción que no necesariamente se exponen como censura directa, sino como actos sutiles de invisibilización digital.

Siendo así, la problemática no recae únicamente en la posibilidad de emitir un discurso, sino en las condiciones reales para que este pueda ser tanto visible como accesible para la comunidad en el debate político. En este sentido, los discursos minoritarios, que suelen ser expresados por comunidades que históricamente han tenido menor poder simbólico, se deben enfrentar con barreras estructurales que limitan su visibilidad en el ecosistema digital pues su contenido es catalogado como conductas “inapropiadas” o sencillamente se elimina (Wells, 2024).

Teniendo en cuenta lo anterior, según IFEX (2025) los defensores de derechos humanos, periodistas o activistas “son especialmente afectados, porque sus relatos son sistemáticamente considerados como «sensibles» o «peligrosos» por sistemas que no entienden el contexto ni el propósito de las expresiones detectadas por los filtros automatizados”. En consecuencia, esto genera el planteamiento de que existen grandes desafíos tanto para la equidad digital<sup>3</sup> como para el pluralismo democrático.

A partir de lo expuesto, el presente texto investigativo analizará el papel que tienen los algoritmos en la configuración del discurso público digital y su incidencia en la invisibilización de grupos minoritarios. Para esto, se adoptará un enfoque crítico que articulará perspectivas tecnológicas, jurídicas y comunicativas, con el propósito de entender las consecuencias de este fenómeno en el ejercicio de derechos fundamentales.

---

<sup>2</sup> Es el proceso mediante el cual los algoritmos informáticos filtran, priorizan y dan forma a nuestras experiencias, especialmente en línea, impactando en las elecciones y percepciones. (Sustainability Directory, s. f.)

<sup>3</sup> La equidad digital se refiere al estado donde cada persona y comunidad tiene los recursos necesarios de tecnología de la información para participar plenamente en la sociedad, la democracia y la economía. El término abarca no solo el acceso a dispositivos e Internet, sino también la capacidad de usarlos y beneficiarse de ellos de manera efectiva. (Internet Society Foundation, 2023)

Es por esto que, para mayor claridad este artículo se estructura de la siguiente manera: en esta sección introductoria se encuentra i) la definición del problema, ii) el propósito de la investigación, iii) el marco teórico y iv) los antecedentes. En la segunda sección se presenta la metodología utilizada; en la tercera sección se desarrollan los principales argumentos conceptuales y empíricos; la cuarta sección expone los resultados sistematizados de la investigación; la quinta ofrece una discusión crítica y propuestas de reforma normativa; y la sexta presenta las conclusiones.

### **3.1 Definición del Problema**

El problema central que aborda este artículo es el silenciamiento digital mediado por algoritmos, en las distintas plataformas digitales a grupos minoritarios. Pues, a diferencia de las formas tradicionales de censura, esta forma de invisibilización opera de manera indirecta y casi imperceptible, bien sea por medio de mecanismos de filtrado y priorización de contenido que llegan a limitar el alcance de ciertos discursos (Civil Liberties Union for Europe, 2021).

Pues bien, es necesario definir que los algoritmos de recomendación “son un conjunto de instrucciones que se ejecutan de manera secuencial, con el fin de proporcionar soluciones de optimización y de esta manera filtrar a los usuarios toda la información de valor o interés” (Korus et al., 2021). Es decir, estos son diseñados con el objetivo de maximizar la interacción y el tiempo de permanencia de las personas, por tanto, tienden a priorizar contenidos que ya poseen altos niveles de popularidad. En consecuencia, este fenómeno genera una amplificación de los discursos dominantes, mientras que aquellos que no pueden insertarse en esas dinámicas, se mantienen en posiciones marginales.

Adicionalmente, los sistemas de Inteligencia Artificial (IA) pueden traer consigo sesgos derivados de los datos con los que son entrenados<sup>4</sup> (IBM s.f), los cuales siguen reproduciendo y reflejando desigualdades históricas y sociales. En consecuencia, ciertos grupos (como minorías étnicas, comunidades LGBTQ+, o actores políticos disidentes) pueden verse injustificada y desproporcionadamente afectados por dinámicas de silenciamiento digital. De hecho, según IFEX (2025) “las experiencias de personas refugiadas y/o racializadas, por ejemplo, se vuelven aún más invisibilizadas”, lo que pone de presente el carácter estructural del problema y su relación con desigualdades sociales más amplias.

En consecuencia, esta problemática no se limita únicamente a la existencia de los algoritmos, sino además en la forma en que estos determinan qué grupos tienen visibilidad en el ecosistema

---

<sup>4</sup> “El sesgo de la IA, también llamado sesgo de machine learning o sesgo algorítmico, se refiere a la aparición de resultados sesgados debido a los sesgos humanos que sesgan los datos de entrenamiento originales o el algoritmo de IA, lo que provoca resultados distorsionados y resultados potencialmente dañinos. (...) Los modelos en los que se basan los esfuerzos de IA absorben los sesgos de la sociedad que pueden estar silenciosamente incrustados en las montañas de datos con los que se entrenan. La recopilación de datos con sesgos históricos que reflejen la desigualdad social puede resultar perjudicial para los grupos históricamente marginados en casos de uso como la contratación, la policía, la calificación crediticia y muchos otros”. (IBM, s.f.)

digital. Sobre todo, aquellos que históricamente han sido marginados. Además, de analizar como en el ecosistema digital, se refuerzan asimetrías en la distribución del poder simbólico<sup>5</sup>.

### **3.2 Propósito de investigación e hipótesis**

El propósito del artículo es ofrecer un análisis crítico sobre cómo los algoritmos, principalmente los de las redes sociales, que rigen las plataformas digitales contribuyen al silenciamiento de los grupos minoritarios. Esto último, detectando las dimensiones económicas, sociales, políticas y técnicas que los conforman. Además, se pretende contribuir al debate académico acerca de la necesidad de una mayor transparencia en el diseño algorítmico, a través de una comprensión más profunda de las dinámicas de poder en el entorno digital.

De la misma manera, se hace imperativo exponer aspectos para una reflexión crítica sobre el papel que juega la tecnología en la construcción del mundo digital. Por tanto, de acuerdo con lo expuesto, las preguntas que guían la investigación son:

- ¿A través de qué mecanismos los algoritmos de moderación de contenidos generan efectos de silenciamiento sobre los discursos de las minorías?
- ¿Qué marcos regulatorios son apropiados para garantizar la igualdad de acceso a la esfera pública digital y el derecho a la libertad de expresión?
- ¿Cuáles son las consecuencias sociales, políticas y culturales de la invisibilidad digital?

Ahora bien, la hipótesis propuesta es que los algoritmos de moderación de contenido, al ser adiestrados principalmente con datos que los usuarios ofrecen (los cuales en muchas ocasiones exponen patrones de discriminación), mantienen su reproducción. Esto, alimenta una forma de discriminación estructural indirecta que llega a vulnerar diversos derechos como: la libertad de expresión, la igualdad y la no discriminación. Por tanto, se requiere de acciones jurídicas que no se limiten a la autorregulación voluntaria de las plataformas.

### **3.3 Marco teórico**

El estudio del silenciamiento digital mediado por algoritmos resulta de un estudio interdisciplinario. De esta manera, para el presente artículo se tendrá como base diversos textos académicos con el objetivo de realizar el análisis correspondiente. Siendo así, en primer lugar, Gillespie (2014) ideó el concepto de “política de las plataformas” para detallar cómo las decisiones de los algoritmos y diseños tecnológicos son intrínsecamente políticas, pese a presentarse como técnicas. En segundo lugar, Pasquale (2015) abordó el concepto de «sociedad de caja negra» (black box society) para exponer que aquellos sistemas operan de forma invisible, lo que ocasiona una incidencia trascendental en ámbitos vitales de la vida de las personas. Esto,

---

<sup>5</sup>El poder simbólico se refiere a la capacidad de usar símbolos, lenguaje y significados para ejercer influencia social y control. Pierre Bourdieu, un sociólogo francés, popularizó este concepto, destacando cómo el poder simbólico legitima y perpetúa estructuras de dominación. Recordar esto puede ayudarte a entender mejor cómo funcionan las relaciones de poder en la sociedad. (StudySmarter, s. f.)

sin permitir que aquellos afectados puedan conocer su funcionamiento y poder impugnarlos o contrarrestar sus efectos.

Adicionalmente, la obra de Safiya Umoja Noble, *Algorithms of Oppression* (2018), demostró cómo los algoritmos de búsqueda de Google perpetúan estereotipos racistas y sexistas, dando mayor exposición a representaciones degradantes de mujeres negras. Esta investigación abrió una línea de trabajo que conecta la arquitectura técnica con la reproducción de la desigualdad racial, inaugurando lo que Benjamin (2019) denominó la «nueva codificación de Jim Crow» en referencia a las leyes de segregación racial del sur de Estados Unidos. Así mismo, Cathy O'Neil (2016), matemática y analista de datos, extendió el análisis al demostrar que los modelos predictivos utilizados en ámbitos como el crédito financiero, la justicia penal y el empleo actuaban como «armas de destrucción matemática» que penalizan sistemáticamente a las comunidades más vulnerables.

Además, en este ámbito también será posible de relacionar con el concepto de “sociedades de control” que propone Gilles Deleuze en su artículo “Postscript on the Societies of Control”. Este autor hace la explicación del tránsito que se evidenció desde las sociedades disciplinadas a las sociedades de control, como lo es ahora. Con esta última, se destaca entonces el rol o la posición en la que se encuentran las plataformas digitales y los mecanismos informáticos. Así, a través de este se determina que hoy en día las personas se representan mayormente a través de códigos y datos. Mientras que en el mundo disciplinado se requiere la presencialidad para realizar cualquier tipo de actividad, en la sociedad de control existen los códigos y los circuitos abiertos que describe Deleuze, con lo cual “se permite una dominación más continua y flexible a través de la modulación y el rastreo de datos en todos los aspectos de la vida, como la salud, finanzas, etc.”(Deleuze, G. 1992)

Con ello, el ser humano se transforma y se aparta de lo que era una sociedad disciplinada, en la cual se “encerraba” en la vida cotidiana y sus actividades para subsistir. (Deleuze,G. 1992) Ahora, todo eso puede hacerlo mediante la tecnología, con lo cual se pierde la noción del “encierro”, un encierro hipotético en el hogar, el trabajo, la esfera de la vida, que pasó a considerarse algo inútil para pasar a encontrar ese sentido en las plataformas digitales. De esto se puede entonces afirmar que, actualmente, todo funciona a través de sistemas digitales automatizados, que las personas ahora son vistas como meros datos personales los cuales pueden ser objeto de control y estudio constante por parte de las instituciones que los manejan, y que se ha convertido en una sociedad dependiente de la esfera digital que se aparta de su funcionamiento tradicional disciplinado.

Pues bien, en concordancia con lo expuesto, es necesario abordar el concepto de discriminación indirecta. Este fue desarrollado inicialmente en el ámbito del derecho laboral anglosajón y posteriormente incorporado al derecho internacional de los derechos humanos. Según la Convención Interamericana contra toda forma de discriminación e intolerancia esta se produce “(...) cuando una disposición, un criterio o una práctica, aparentemente neutro es susceptible de

implicar una desventaja particular para las personas que pertenecen a un grupo específico (...)” (OEA, 2013). Es decir, se designa a aquellas prácticas, normas o criterios que parecen ostentar un carácter de neutralidad, pero que, en la práctica, producen efectos desproporcionadamente nocivos a un grupo protegido. El Tribunal Europeo de Derechos Humanos ha consolidado esta doctrina en numerosas sentencias, entre las que destaca D.H. y otros c. República Checa (2007), en la que el Tribunal estableció que la discriminación indirecta puede derivar de medidas que, sin perseguir un fin discriminatorio, producen en la práctica ese resultado (Tribunal Europeo de Derechos Humanos, 2007).

Siguiendo con lo expuesto, la inserción de esta doctrina al campo de la moderación algorítmica resulta conceptualmente posible, sin embargo, tiene grandes dificultades jurídicas. Siendo así, los sistemas automatizados de filtrado y clasificación de contenido no poseen de manera explícita categorías discriminatorias. Empero, como exponen los estudios de Buolamwini y Gebru (2018) sobre los sistemas de reconocimiento facial, la imprecisión diferencial de estos sistemas según variables de raza y género, así como el entrenamiento sobre datos históricos sesgados, producen resultados que afectan desproporcionadamente a colectivos vulnerables. En el ámbito específico de la moderación de contenido, investigaciones de Derechos Digitales América Latina (2025) han documentado sistemáticamente cómo los sistemas automatizados generan fallos en la amplificación del contenido producido en lenguas minoritarias, dialectos, jergas comunitarias y formas de expresión culturalmente situadas, invisibilizando discursos legítimos de resistencia y denuncia.

Ahora bien, frente a la esfera pública digital, la teoría habermasiana de la acción comunicativa (Habermas, 1981) y sus desarrollos posteriores en torno al concepto de esfera pública (Habermas, 1989) proveen un marco normativo de referencia para comprender las implicaciones democráticas del silenciamiento digital. La esfera pública, entendida como el espacio social en el que los ciudadanos deliberan racionalmente sobre los asuntos de interés general, requiere, para su funcionamiento adecuado, condiciones de acceso igualitario, pluralismo de perspectivas y ausencia de mecanismos de exclusión basados en el origen o la posición social de los participantes.

Adicionalmente, las plataformas digitales si bien en principio ofrecen una ampliación sin precedentes del acceso a la expresión pública (reduciendo los costos de entrada que caracterizan a los medios de comunicación tradicionales), su arquitectura técnica introduce nuevas formas de estratificación y jerarquización del discurso. El concepto de «arquitectura de la elección» (Thaler & Sunstein, 2008), aplicado al diseño de interfaces y algoritmos, ilustra cómo las decisiones técnicas moldean de manera imperceptible los comportamientos y las posibilidades discursivas de los usuarios. Zuboff (2019) añade una dimensión adicional al describir el «capitalismo de vigilancia» como un nuevo régimen económico que extrae valor de la conducta humana, lo que implica que los algoritmos de las plataformas no buscan primariamente la verdad o la justicia comunicativa, sino la maximización del engagement y, con ello, la rentabilidad publicitaria.

Por último, surge el planteamiento de los derechos digitales como derechos humanos. La evolución del derecho internacional de los derechos humanos ha reconocido progresivamente que los derechos fundamentales no se agotan en el espacio analógico, sino que se proyectan con plena vigencia sobre el entorno digital. La Resolución 20/8 del Consejo de Derechos Humanos de las Naciones Unidas (2012) estableció que “los mismos derechos que tienen las personas fuera de línea deben estar también protegidos en internet”, sentando las bases de lo que se ha denominado el principio de equivalencia digital. El relator especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, Frank La Rue, profundizó en este marco en su informe de 2011 (A/HRC/17/27), en el que identificó el acceso a internet como un derecho habilitante de múltiples dimensiones de la libertad de expresión.

En el contexto interamericano, la Comisión Interamericana de Derechos Humanos (CIDH) y su Relatoría Especial para la Libertad de Expresión han desarrollado una doctrina progresiva sobre los estándares aplicables a la expresión digital. La Declaración Conjunta sobre Libertad de Expresión e Internet de 2011, adoptada por los relatores especiales de la ONU, la OSCE, la OEA y la CADHP, estableció que los intermediarios de internet no deben ser responsables por el contenido de terceros, salvo que hayan contribuido activamente a él o no hayan actuado con rapidez para retirarlo tras ser informados de su carácter ilegal. Este marco, sin embargo, no contemplaba suficientemente los supuestos de moderación algorítmica proactiva ni sus efectos diferenciados sobre comunidades vulnerables.

## **IV. METODOLOGÍA**

### **4.1 Diseño del estudio:**

La investigación de este artículo sigue las bases de un método de análisis cualitativo, basado en documentación de carácter descriptivo, analítico e interpretativo, enfocado en el entendimiento de las dinámicas sociales y jurídicas complejas aplicadas a este enfoque moderno. Este método se implementa bajo el entendido de que responde perfectamente al objeto del presente artículo: el silenciamiento de los discursos minoritarios a través de algoritmos, tema que no presta para determinar por medio de una estructura cuantitativa.

La escogencia de este enfoque responde a la naturaleza del objeto de estudio, el cual resulta imposible observar, interpretar, y por consiguiente concluir, exclusivamente mediante herramientas cuantitativas. Así, la metodología cualitativa permite, por un lado, abordar este complejo fenómeno desde múltiples perspectivas, estudios y opiniones, y por otro, otorgar una interpretación aún más profunda de quien lo estudia.

De esta manera, el diseño metodológico se sustenta en una revisión doctrinal sistemática, que incluye fuentes de distinta naturaleza: textos académicos, doctrina de organizaciones internacionales, marcos normativos (nacionales e internacionales), documentos institucionales,

entre otros. Siendo así, esta pluralidad de fuentes permite abordar el fenómeno de silenciamiento digital desde múltiples aristas, permitiendo una comprensión integral del problema.

#### **4.2 Método de análisis:**

Una vez recopilada la información necesaria para este artículo, se empleó la siguiente estructura de análisis. En primera medida, se llevó a cabo una lectura crítica de cada uno de los documentos escogidos, resultando en una categorización y definición de toda la evidencia tanto empírica como jurídica de lo recaudado en el marco de lo que aquí concierne, discursos minoritarios que son invisibilizados.

En segundo lugar, se ejecutó una estructura comparativa, tanto de marcos normativos (externos e internos), como jurisprudencia y opiniones de todo tipo, con el objetivo principal de contrastar todos los modelos y aristas que puede tener este asunto. Así, consistió en un ejercicio de identificación y determinación de aspectos similares, diferencias y hasta vacíos en la manera en cómo se aborda la problemática de silenciamiento digital.

Por último, este artículo busca plantear una posición interpretativa derivada de toda la información encontrada, que permite también adoptar una posición crítica, aunque al mismo tiempo objetiva, de este fenómeno con el fin último de profundizar en esta materia, de la que poco se investiga. De esta manera, el presente texto, no se limita a describir el fenómeno, sino que además busca aportar aspectos para su comprensión y debate académico.

#### **4.3 Limitaciones metodológicas encontradas:**

En el transcurso de esta investigación, fue posible determinar la existencia de varios problemas. De hecho, existió el surgimiento de dudas imposibles de resolver, de acuerdo al tema aquí analizado. Por un lado, la velocidad inminente de crecimiento, fortalecimiento y adaptación de los mecanismos tecnológicos. Como es sabido, la digitalización algorítmica avanza de forma rápida, siendo así, es poco probable que las mismas prácticas se mantengan en el tiempo. Razón por la cual, se torna difícil la durabilidad de este análisis toda vez que pueden emplearse diferentes herramientas y van surgiendo innovaciones digitales que tornan este tipo de estudios mucho más complejo.

Por otro lado, las plataformas tecnológicas emplean estructuras complejas de descifrar y comprender cuyo funcionamiento interno impide a la comunidad acceder a un análisis preciso y transparente del funcionamiento de este tipo de mecanismos. Con ello, su operación se convierte en tedioso asunto para interpretar.

## V. ARGUMENTACIÓN

### 5.1 El algoritmo

Las estructuras de visualización de contenido, hoy en día conocidas como las grandes y reconocidas plataformas: Instagram, Tik Tok, Youtube, Facebook, entre otras, son populares por ser nuevas alternativas de acceso a la información. Estas funcionan a través de un método llamado el algoritmo, encargado de distribuir y gobernar el contenido al cual acceden los distintos usuarios. La Universidad Europea lo define como “una secuencia ordenada y finita” (Universidad Europea, 2025). Sin embargo, a pesar de tener sentido esta definición, en un ámbito digital va mucho más allá.

El reconocido diario español “Comunicados”, hace referencia al algoritmo como “ese conjunto de reglas y flujos en el marco de la programación que decide si se muestra una publicación o no” (Comunica2, 2021). Ahora bien, en la práctica, realmente no sólo decide qué información circula o no, sino también cuándo y cómo. Con esto, es sencillo precisar que las herramientas a través de las cuales funcionan las plataformas digitales no son para nada neutrales ni objetivas. Por el contrario, se torna evidente que consisten en entidades creadas y destinadas a reproducir, por un lado estructuras de poder, y por el otro un amplio repertorio de patrones de desigualdad visibles en la sociedad.

Overbye-Thompson, H., & Rice, R.E., indican que a raíz de los algoritmos surge algo llamado el “sesgo algorítmico”. Estos autores destacan en un conjunto de etapas a raíz de las cuales el sesgo del algoritmo modifica el desarrollo humano, indicando principalmente “(1) datos sesgados o manipulados, (2) personas sesgadas al tomar decisiones y (3) creación de más estructuras algorítmicas sesgadas” (Overbye-Thompson & Rice, 2025). Con ello, se puede establecer que el sesgo algorítmico es algo que se ha impuesto, más no algo que surja de manera natural, para obtener ese control sobre la sociedad y su respectivo acceso a la información, así como también para modificar la manera en la que interpreta y analiza los diferentes acontecimientos sociales, culturales y económicos de cara a comunidades minoritarias.

De lo anterior, nace una preocupación latente en el desarrollo humano, pues además de originarse un sesgo en la forma en la que los usuarios reciben la información a través de plataformas digitales, eso también resulta en un cúmulo de formación de sesgos. Conforme esto avanza, el mismo algoritmo se torna sesgado, lo que termina por producir un efecto dominó en la forma de desarrollar el pensamiento crítico y auténtico de la humanidad.

Ahora bien, Lee, Resnick y Barton han estudiado la manera en la que estos algoritmos incentivan un violento capitalismo, reflejando las diferencias entre clases sociales, grupos mayoritarios y grupos minoritarios. Así, estos autores precisan que el algoritmo funciona de tal manera que “muestran lo que el usuario quiere ver” (Barton, G., Lee, N. T., & Resnick, P., 2019). Sin embargo, lo que reflejan en realidad es “una perspectiva desproporcionada del lenguaje y de

múltiples patrones de relacionamiento entre grupos sociales, con lo cual se observa implícitamente la clasificación de - anómalo, irrelevante, malo, dañino - al contenido que proviene de comunidades que no se ajustan necesariamente a esos modelos dominantes de otros grupos mayoritarios” (Barton, G., Lee, N. T., & Resnick, P., 2019)

Estos mismos autores exponen un ejemplo casuístico hipotético, el cual demuestra la práctica poco objetiva del algoritmo, que, alimentado por otros sesgos sociales, reproduce esta modalidad de discriminación contra grupos minoritarios. Así, hacen alusión a que, como es común actualmente, “si un afroamericano tiene más probabilidad de ser arrestado en los Estados Unidos de América debido al racismo histórico y desigualdades dentro del sistema social y judicial, estas realidades se reflejarán de igual manera en los algoritmos, o como ellos lo llaman datos de entrenamiento, y se utilizan para reforzar aquella posición” (Barton, G., Lee, N. T., & Resnick, P., 2019) De esto es posible inferir que, el algoritmo viene sesgado en sí mismo, pues al incorporarse en él este tipo de sucesos históricos orientados a reforzar la discriminación de minorías, este solo se encargará de emitir el mismo tipo de juicio a través de la información que muestra a los usuarios.

En concordancia con lo descrito, es viable establecer que esta dinámica alimenta una teoría denominada el “digital Matthew effect”. Como se ha visto, esta dinámica implementada por el algoritmo solo enaltece de cierta manera a los grupos mayormente aceptados, con dinámicas socialmente “comunes”. Así, Overbye-Thompson, H., & Rice, R. E. citan en su artículo al autor Merton, quien describe el “digital Mathew effect” como esa idea que refuerza el hecho de que “los grupos sobrerrepresentados en este tipo de datos de entrenamiento (lo que se llamaría algoritmos actualmente) en las diferentes funcionalidades del sistema, como lo sería hoy en día las plataformas digitales, obtienen ventajas y beneficios adicionales, mientras que comunidades marginadas ven reflejada su propia exclusión y discriminación sistemática.” (Overbye-Thompson & Rice, 2025)

Con esto, es evidente que el algoritmo, basándose en estas bases socialmente creadas sin fundamento, solo refuerza la invisibilidad de las minorías sociales, haciéndolos a un lado en contextos tan influyentes como lo son las redes sociales. En ese mismo sentido, resulta adecuado profundizar en el segundo punto de análisis, descrito como la falsa neutralidad algorítmica.

## **5.2. Neutralidad algorítmica inexistente**

El origen de este tipo de ideas puede verse, por ejemplo, en la teoría de Shoshana Zuboff, quien profundiza en el concepto del “capitalismo de vigilancia”. Este autor lo considera como la definición perfecta de la nueva etapa en la que se encuentra el mundo entero. Este concepto lo define específicamente como “una lógica económica parasitaria en la que la producción de bienes y servicios se subordina a una nueva arquitectura global de modificación del comportamiento”(Zuboff, S. 2019) o como “el origen de un nuevo poder que ejercen dominio sobre la sociedad y plantea desafíos alarmantes.”(Zuboff, S. 2019) Así, bajo este concepto,

podría estarse ante un escenario en el que la experiencia humana pasa a convertirse en insumo con lo cual controlar a la sociedad.

De lo anterior, el autor precisa que “el capitalismo de vigilancia reclama unilateralmente la experiencia humana como materia prima gratuita para su traducción en datos de comportamiento.”(Zuboff, S. 2019) Con esto, se refuerza la idea de que las plataformas digitales podrán ser aquel instrumento a través del cual se ejerce un control por parte de quienes controlan los datos, la visibilidad y los algoritmos junto con la experiencia de cada individuo.

Así, lo que se piensa que es simple contenido, moda, tendencia o visualización, en el fondo es una estructura perfectamente diseñada para modificar la mentalidad de quien está viendo. Razón por la cual, aquellas plataformas terminan por definir lo que ellas quieren que el público vea, pues ostentan el poder y la “soberanía” de sus propias plataformas. Esto termina por ser hasta ingenioso, pues no requiere de una coerción hacia el usuario explícita, por el contrario, se torna muy sencillo.

El mismo autor hace énfasis en un asunto muy curioso, los “productos predictivos”. Este tipo de estructuras analizan el comportamiento de tal manera que son capaces de predecirlo. El mundo digital se ha vuelto tan avanzado, haciéndole creer a la población que hasta es escuchada por los mismos dispositivos móviles. Por un lado, por supuesto que no se niega su utilidad toda vez que ha traído gran ayuda y eficiencia, pero por otro lado se ha excedido su capacidad. Esto se puede observar conforme lo establece el mismo autor mencionado, que a través de aquellos datos de comportamiento que se generan por los usuarios, “estos se incorporan a procesos de fabricación avanzados conocidos como inteligencia artificial y se transforma en productos predictivos que anticipan lo que harás ahora, pronto y en el futuro.”(Zuboff, S. 2019)

La forma en la que funciona el control humano a través de redes sociales y plataformas digitales con sus algoritmos también puede relacionarse y verse como el concepto del Panoptismo que menciona Michael Foucault en “Discipline and Punish: The Birth of the Prison.” Este libro introduce el concepto del Panoptismo basado en el diseño de la prisión ideal, el panóptico, de Jeremy Betham. Con esto, es preciso definir que “el panóptico es un anillo de edificios que rodea una torre central a través de la cual, debido a la posición de las ventanas y la luz, el vigilante de esta estructura arquitectónica puede ver en todo tiempo y lugar a los detenidos de la prisión, sin que estos tengan certeza de si están siendo observados o no.”(Foucault, M. 1995) Este es el tipo de control que se ejerce a través de los algoritmos, formando así su dictadura, mediante mecanismos digitales; un control pasivo, tal vez no constante pero presente, en donde su destinatario no es consciente de ser observado pero que si piensa en esa probabilidad, lo más probable es que cambie su comportamiento.

De lo anterior se busca hacer la analogía de lo que sería el efecto de la mirada invisible. Es decir, se puede observar como la teoría de la torre en la prisión de Foucault se puede evidenciar también en el funcionamiento actual de las plataformas digitales y redes sociales. En la teoría de

Foucault se busca “inducir en el detenido un estado consciente y permanente de visibilidad que garantiza el funcionamiento automático del poder”(Foucault, M. 1995), con el cual, a través del modelo arquitectónico de la torre, no es necesario que esa vigilancia se ejerza por completo y de manera efectiva y constante, por el contrario, solo basta con que el prisionero cuente con la certidumbre de que podría estar siendo observado. Así, el poder que se ejerce no está siendo directo, sino que se ejerce a través del simple mecanismo de la arquitectura. Esto es lo mismo que sucede con la tecnología, pues se analiza el comportamiento del usuario viéndolo y observando no directamente, sino a través de la estructura que en este caso será la plataforma digital.

Con ello, el conglomerado de entornos digitales ya sea plataformas virtuales, redes sociales, bases de datos, entre otras, serán la estructura, como lo es para Foucault la torre, que se encargará de ejercer ese mecanismo de control pasivo. Así, el usuario no solo se siente observado por su plataforma digital o dispositivo móvil, el cual a través de su algoritmo sabe lo que quiere o desea, sino que su propia sujeción de considerar que es observado también cambiará su comportamiento, su decisión y su propósito en un entorno digital. Ello lo guiará a hacer otra cosa distinta a la que en principio quería hacer porque está esa mirada invisible que lo hace sentirse vigilado constantemente, a pesar de que directamente no se sienta observado. Es lo que hace el algoritmo, ejercer un control pasivo observando el comportamiento con el fin de que sea modificado.

Sobre esto resulta importante agregar que, esta clase de algoritmos son creados a partir de un mecanismo previamente pensado y determinado. Es decir, se puede inferir que estos se entrenan, y se entrenan a partir de conjuntos de datos que conocen sobre la marcha, con lo cual no son estructuras simples que solo distribuyen información de forma aleatoria, sino que son producto de prácticas humanas preestablecidas, las cuales resultan ser estructuralmente desiguales.

Adicionalmente, O'neil trae un escenario de discriminación en su libro titulado “Weapons of math destruction: How big data increases inequality and threatens democracy”, que cuando, por ejemplo, “decisiones favorables a blancos sobre negros en materia de privación de la libertad” (O'Neil, 2016), y esto se puede traer al escenario del algoritmo, pues es preciso concluir que funciona como una sistema de este tipo; aprende de datos históricos, refleja el mismo concepto discriminatorio, y lo replica fácilmente. Conforme a la teoría de O'neil, estas prácticas se vuelven “tóxicas, de tal manera que empiezan a desarrollarse como ciclos repetitivos que ayudan a sostener estos discursos.” (O'Neil, 2016) Ese es el problema al que se enfrenta la humanidad con los algoritmos en redes sociales, que perpetua esa discriminación que se da en el ejemplo a través de contenido que decide circular.

Ahora bien, otro aspecto fundamental relacionado a este punto de análisis es el reconocido shadow banning, que más adelante se abordará con mayor profundidad. Este concepto, es definido por Mauro, Emiliano, Alessandro, Pujan y Gianluca, como “esa posibilidad del algoritmo de limitar y restringir cierto contenido.” (Conti, M., Cristofaro, E., Galeazzi, A.,

Paudel, P., & Stringhini, G., 2024) Igualmente, resaltan que los algoritmos “están diseñados para propender a minimizar los debates en las redes sociales, más bien lo que buscan es difundir el contenido más fiable y saludable posible para fomentar un ecosistema menos agresivo”. (Conti, M., Cristofaro, E., Galeazzi, A., Paudel, P., & Stringhini, G., 2024) Sin embargo, de esto surge la gran discusión, pues es necesario también difundir sucesos en donde se afectan los derechos de comunidades minoritarias, con el fin de elevarlas en la sociedad.

Comúnmente, cuando ocurre algo hacia una comunidad marginada, sea bueno o malo, no es visible en las redes. En ocasiones, estos asuntos se tornan de especial importancia toda vez que puede ser necesaria la circulación de una vulneración de derechos humanos a comunidades marginadas, lo cual requiere que el mundo visualice y, a su vez, alcance a los entes especializados necesarios para atender dichas situaciones. Sin embargo, ahí es cuando entra el algoritmo a invisibilizar aquellas situaciones, discursos, afectaciones, hasta opiniones de estos grupos, que los apartan del mundo “real”, dejándolos en un espectro que ningún juez alcanza.

Partiendo de lo aquí expuesto, el algoritmo que ya contiene en sí mismo el propio sesgo, no solo es capaz de predecir lo que el usuario quiere consumir, o lo que va a consumir, sino que igualmente es capaz de predecir la idea que creará en la cabeza del usuario, que consiste en ese concepto reforzado de discriminación y silenciamiento de los grupos marginados de la sociedad.

Es decir, no solo circulará a su antojo el contenido que considere “relevante” o “apropiado” que no “genera debate o violencia en redes sociales”, sino que permea el conocimiento del usuario, lo viciado del mismo sesgo como un círculo igual de vicioso. Ello hace cada vez menos visible y accesible el discurso de otros, de manera desventajosa puestos en sociedad, alimentando esa desigualdad. Así, se hace aún más fuerte la hipótesis de que el algoritmo no es nada inofensivo, mucho menos neutral, por el contrario, se refuerza la idea de que son mecanismos de control de masas, incluyendo en comportamientos e ideas que resultan erróneas en la población.

### **5.3. El algoritmo y la libertad de expresión**

Actualmente en Colombia, y también de manera internacional, se ha tratado de otorgar una utilización de estos modelos tecnológicos muy garantistas. Si bien, la Corte Constitucional define los espacios de entornos digitales y redes sociales como “un lugar a través del cual se da un intercambio de ideas libre y masivo, donde las personas son capaces de expresar opiniones y posturas que llegarán a miles de personas” (Sentencia T-256, 2025. C.C.), la Corte también reconoce que este es un escenario que se presta para la vulneración de derechos masiva. En principio, la confrontación que nace sobre el libre desarrollo de la personalidad y libertad de expresión, pues las redes “amplifican los sesgos, los odios y la violencia del mundo” (Sentencia T-256, 2025. C.C.), así como también se presentan afectaciones a los derechos de habeas data, el debido proceso, autonomía de la voluntad, entre otros.

Sin embargo, a pesar de que el uso de redes sociales puede chocar con los derechos como el de la libertad de expresión, la Corte Constitucional por ejemplo aterriza un poco más el alcance del mismo en Colombia y el mundo digital, recordando que ningún derecho es absoluto. Así, teniendo en cuenta que la libertad de expresión significa que “todas las personas pueden expresar sus pensamientos a través de opiniones y de difundir información, y que la población es capaz de conocer las ideas de otros” (Sentencia T-256, 2025. C.C.), está limitado en tanto que choque con el derecho al buen nombre y la honra de otra persona. Igualmente, la Corte precisa sobre discursos que se encuentran completamente prohibidos como por ejemplo propaganda a favor de la guerra, incitación directa y pública a cometer genocidio, apología de odio nacional, racial o religioso que incite a la discriminación, violencia y demás. Esto debido a que pueden generarse y materializarse actos violentos en contra de poblaciones vulnerables, minorías o hasta personas.

De igual manera, la jurisprudencia de la Corte es enfática en decir que la libertad de expresión debe ser garantizada también en espacios digitales, pues indica que este derecho también “cobija las manifestaciones ofensivas, chocantes, impactantes, indecentes, escandalosas, excéntricas o simplemente contrarias a las creencias y posturas mayoritarias.” (Sentencia T-256, 2025. C.C.) Ello se vio analizado en la emblemática sentencia T-256 de 2025 en donde Esperanza Gómez Silva alegaba la protección de este derecho a pesar de ejercer la profesión de creación de contenido para adultos. En tal sentido, cuando se habla por ejemplo del mecanismo de moderación de contenido, si se remueve o elimina un contenido específico se torna necesario verificar a profundidad que este “obedezca exclusivamente al tono de las publicaciones o al hecho de que no corresponda a una postura mayoritaria” (Sentencia T-256, 2025. C.C.) y tratarse de buena forma, pues podría estarse ante una posible vulneración del derecho de la libertad de expresión.

Con ello, se desprende un debate sobre el control que se ejerce a través de estas plataformas mediante la moderación de contenido, pues este se ha convertido en el eje central de las redes sociales que a su misma vez se torna en un mecanismo de control algorítmico a los usuarios. Esto se vuelve un aspecto muy relevante toda vez que se contrapone con otros derechos como el debido proceso. Por un lado, según la Corte Constitucional, la moderación de contenido consiste en “la práctica organizada de revisión de los contenidos generados por usuarios en páginas de internet”, (Sentencia T-256, 2025. C.C.) la cual faculta a la plataforma a eliminar el contenido que considere contrario a las normas comunitarias. Podría llamarsele un límite a la libertad de expresión, pues finalmente las plataformas deciden que puede o quieren que circule en las redes y que no.

Ahora bien, el derecho a la libertad de expresión también está consagrado en normativas pertenecientes a otros países, incluyendo diversos instrumentos internacionales, propios del derecho internacional. Como si de manera conjunta, el mundo entero considerara este derecho como algo de carácter tan fundamental, que es necesario preservar en todo el globo. Así, mecanismos como la Declaración Universal de Derechos Humanos, el Pacto Internacional de Derechos Civiles y Políticos, el Convenio Europeo de Derechos Humanos, la Carta de los

Derechos Fundamentales de la Unión Europea, hasta la Constitución Española, por ejemplo, dan fe y constituyen prueba fiel de que esta garantía de libertad no es un tema de menor relevancia.

Con ello, tal y como lo describe Calcaneo, las redes sociales demuestran por un lado el gran vacío que existe de cara a la protección del derecho de la libre expresión, y por otro lado, la ausencia de regulación de los entes del estado respecto de este tema en escenarios tan complejos como estos. Así, este autor indica que, por ejemplo, “actualmente derechos como la libertad de expresión coexisten en un modelo privado de regulación aplicada únicamente al Internet”. (Calcáneo, 2021) En efecto, tiene razón, ya no representan dos escenarios estrictamente apartados y diferentes, al contrario, conviven y se alimentan mutuamente.

Asimismo, Calcaneo también afirma que “el sentido regulatorio de este derecho en este entorno digital se comprende de dos asuntos fundamentales; las personas sometidas al mismo a una regulación estatal insuficiente en esta materia y a la regulación propia de las plataformas digitales para el Internet.” (Calcáneo, 2021) Esto, de igual manera, también se reviste de verdad, pues el sistema normativo y regulatorio se queda corto al intentar aplicar el derecho en situaciones inmersas en redes sociales, tanto así que las propias plataformas crearon su propio sistema sancionatorio y permeado de reglas. Con ello, la humanidad se ve sujeta entonces a acoplarse a dos sistemas “judiciales” alternos, con vacíos y con extrema vulneración de derechos. Finalmente, ninguno resulta totalmente eficaz para erradicar desigualdades, situaciones discriminatorias, y menos estereotipos, pues los mismos algoritmos se encargan de replicar contenido de este tipo.

Este mismo autor, cita en su artículo a Mark Zuckerberg, quien recientemente expresó y admitió el poder que tienen las redes sobre el discurso; “plataformas como Facebook, cuentan con un sistema integrado propio de gobernanza que permite determinar el contenido que circula en redes sociales” (Calcáneo, 2021), con lo cual se confirma lo planteado en este artículo, las personas solo ven lo que el control de masas quiere que vean.

De lo anteriormente descrito, a pesar de que en jurisdicciones como Colombia, se hace un intento de regular y proteger el derecho a la libertad de expresión en los entornos digitales, sigue siendo una materia bastante difícil de comprender y cuidar. El espectro digital dificulta la forma en la que conviven derechos como este, en especial cuando los sistemas judiciales preestablecidos mantienen activa la duda de limitar, por un lado, la libertad de expresión, o por el otro, las redes sociales y su alcance algorítmico. Con ello, vale la pena cuestionar hasta qué punto un algoritmo, con todas las características aquí descritas y toda la estructura de su funcionamiento tan bizarro planteada, puede estar desviando los discursos de las minorías, y si esto debe tener más alcance, o cual es el método para tomar de cara al algoritmo per se para combatir este silenciamiento.

## 5.4 Discriminación algorítmica indirecta y el derecho a la igualdad

El fenómeno del silenciamiento digital reviste de gran importancia cuando se estudia a la luz del derecho a la igualdad, en particular teniendo en cuenta la posibilidad de la existencia de una discriminación indirecta. Pues bien, a diferencia de la discriminación directa -que implica un trato diferenciado explícito y deliberado (Salomé, 2017) - la indirecta, es cuando una práctica o criterio que aparenta ser neutral produce efectos desproporcionadamente negativos sobre determinados grupos (Salomé, 2017).

Aunando a lo anterior, desde el plano jurídico internacional, el principio de igualdad y no discriminación ha sido interpretado en un sentido sustantivo, que no se limita a la mera igualdad formal. El Comité de Derechos Humanos de la ONU, en su Observación General N.º 18 (1989), definió la discriminación como:

(...) toda distinción, exclusión, restricción o preferencia que se basen en determinados motivos, como la raza, el color, el sexo, el idioma, la religión, la opinión política o de otra índole, el origen nacional o social, la posición económica, el nacimiento o cualquier otra condición social, y que **tengan por objeto o por resultado** anular o menoscabar el reconocimiento, goce o ejercicio, en condiciones de igualdad, de los derechos humanos y libertades fundamentales. *(subrayado y negrilla fuera del texto original)*

Siendo así, la referencia al “resultado” es central para el estudio del silenciamiento algorítmico. Lo anterior, ya que no basta con tener una intención de discriminar, pues, si los sistemas de recomendación generan efectos desproporcionadamente nocivos sobre grupos colectivos protegidos (que coinciden con ser históricamente marginados), se configura una forma de discriminación indirecta, lo cual es jurídicamente trascendental.

En este mismo sentido, el Sistema Interamericano, reconoce y ha reforzado el carácter estructural del principio de igualdad. De hecho, la Corte Interamericana de Derechos Humanos (Corte IDH), ha reconocido en su Opinión Consultiva OC-18/03 sobre la condición jurídica y los derechos de los migrantes indocumentados que el principio de igualdad y no discriminación pertenece al jus cogens internacional y genera obligaciones erga omnes de carácter inderogable (Corte IDH, 2003). Esto indica que cualquier práctica (incluyendo aquellas mediadas por tecnologías privadas) que produzca efectos discriminatorios debe ser objeto de especial observancia.

Aplicado al entorno digital, y con base en lo ya expuesto, pese a que los algoritmos se exponen como herramientas que permiten la objetividad y son neutrales, no evita que puedan reproducir e incluso amplificar desigualdades estructurales. Siendo así, la priorización de contenidos basados en métricas como el “engagement”<sup>6</sup> o la popularidad, tiende a favorecer discursos dominantes (García, s.f.). Mientras que los discursos minoritarios provenientes de comunidades

---

<sup>6</sup> “Nivel de interacción y participación que genera una marca, contenido o publicación entre su audiencia” (Universidad Europea, 2025)

racializadas, migrantes o disidentes enfrentan mayores limitantes para alcanzar visibilidad (García, s.f.). Como consecuencia, es posible afirmar que el silenciamiento digital se puede categorizar como un ejemplo de discriminación indirecta, en tanto deriva de criterios que parecen ser objetivos, pero pueden llegar a generar resultados excluyentes.

Adicionalmente, en el contexto colombiano, el derecho a la igualdad se encuentra consagrado en el artículo 13 de la Constitución Política de 1991. En ese no se limita a proteger una igualdad formal ante la ley, sino que, además, impone al Estado el deber de promover condiciones para que sea real y efectiva, puede ser por medio de la adopción de medidas en favor de grupos discriminados o marginados. (Constitución Política de Colombia, 1991, art. 13).

Siendo así, la Corte Constitucional ha desarrollado ampliamente este mandato, señalando que la igualdad material exige analizar no solo el trato diferenciado explícito, sino también los efectos desproporcionados que determinadas prácticas pueden generar sobre poblaciones en situación de vulnerabilidad (Sentencia C-038, 2021, C.C). Aplicado al entorno digital, este enfoque implica que los sistemas de moderación y recomendación algorítmica deben ser evaluados no solo por su diseño formal, sino por sus consecuencias reales en la distribución de la visibilidad del contenido. Así, cuando dichos contribuyen a la exclusión sistemática de discursos provenientes de grupos minoritarios, se configura una tensión directa con el principio de igualdad material, lo que exige una revisión crítica desde el derecho constitucional y la adopción de medidas que garanticen condiciones equitativas de participación en la esfera pública digital.

Ahora bien, la violación al principio de igualdad puede generar afectaciones a más garantías, como la vulneración del debido proceso en los sistemas de moderación de contenido. En la práctica, cuando se eliminan algunas publicaciones o hay suspensión de cuentas bajo el presupuesto de que se está vulnerando alguna norma de comportamiento, los mecanismos para impugnar o apelar estas decisiones son absolutamente ineficientes.

En concordancia con lo expuesto, la anterior situación fue evidenciada en un caso revisado por la Corte Constitucional colombiana, en el que se eliminó la cuenta de un usuario bajo la presunción errónea de que era menor de edad, pues se pensaba que tenía 13 años de edad. A pesar de los múltiples intentos del afectado por aclarar su situación y apelar la decisión, la plataforma no atendió sus solicitudes. (Sentencia T-453, 2024. C.C.) Este caso ilustra cómo la moderación automatizada puede derivar en decisiones arbitrarias sin garantías mínimas de defensa, contradicción o revisión. Por tanto, este tipo de herramienta tiene la posibilidad de recaer en una discriminación indirecta que propague la violación constante de derechos fundamentales.

Teniendo en cuenta lo anterior, según Douek (2021) el modelo de moderación de contenido deberá ajustarse a principios de proporcionalidad y necesidad. El primero obedece a que, como lo propone el autor, existen limitaciones injustificadas al derecho de libertad de expresión, pues en ocasiones se está ante un discurso político protegido o meras declaraciones que no violan los derechos de otros y no pueden ser eliminados fácilmente (Douek, E. 2021). Con ello, insiste en que esas limitaciones al derecho son válidas, pero se debe buscar que persigan un fin legítimo,

para que cuando se deba eliminar algún contenido, sea a través del medio menos restrictivo y mantenga una relación (nexo causal) que sea razonable entre el beneficio del límite y el costo para la expresión (Douek, E. 2021). Por su parte, el principio de necesidad esclarece que en este entorno siempre habrá lugar a error, a través del cual podrán evidenciarse falsos positivos y falsos negativos en eliminación de publicaciones o cuentas. Sin embargo, se indica que las plataformas deben buscar la forma más transparente para realizar estos procesos con el fin de minimizar afectaciones injustificadas.

Esta lectura ofrece un concepto muy interesante, la gobernanza digital y la gobernanza del discurso en línea. Esto lo muestra como un nuevo desafío en este mundo digitalizado, en donde las plataformas de redes sociales y la tecnología en general ejerce un poder de gobernanza sobre las personas, el cual se da específicamente en el contexto de la moderación de contenido. Con ello, precisa en que realmente la gobernanza digital es la que evalúa la legalidad o ilegalidad de una publicación, con lo cual es imposible aplicar ese juicio de legalidad a miles de posts diarios, por ende, deberá emplear los mecanismos más eficaces, preferiblemente humanizados, para ejercer su control sobre las publicaciones. (Douek, E. 2021)

En concordancia con lo anterior, es pertinente traer a colación a Dvoskin, quien se pronuncia sobre la alineación de la moderación de contenido con los derechos humanos. Este concepto consiste en que “las redes sociales y las plataformas alineen sus políticas internas de discurso con el derecho internacional de los derechos humanos” (Dvoskin, B. 2023) con el fin de que en este entorno, las empresas actúen como un respectivo sistema de gobernanza legítimo sin que se vulneren derechos fundamentales en las redes sociales, con lo cual se tendrá en cuenta conceptos como el interés público global. Sin embargo, la autora también hace énfasis en que debe ser tratado con mucho cuidado, so pena de convertirse en un modelo que desplace el poder hacia un solo ente determinado respaldado bajo la idea errónea de empoderar a la gente y sus derechos. (Dvoskin, B. 2023)

Pues bien, es menester señalar que la aplicación de esta doctrina al ámbito de la moderación algorítmica plantea el desafío de la causalidad. Es decir, demostrar que los efectos discriminatorios de un algoritmo son consecuencia de su diseño y no de otros factores. Por tanto, los estudios de Buolamwini y Gebru (2018) sobre sesgo en sistemas de reconocimiento facial ofrecen un modelo metodológico para establecer esta causalidad: mediante pruebas controladas, demostraron que los sistemas comerciales de reconocimiento facial de IBM, Microsoft y Face++ tenían tasas de error significativamente más elevadas para mujeres de piel oscura (hasta un 34,7% de error) en comparación con hombres de piel clara (0,3% de error) (Bustelo, 2024) . En consecuencia, esto constituye evidencia de discriminación algorítmica directamente atribuible al diseño del sistema.

Ahora bien, en el marco normativo el Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo relativo a un mercado único de servicios digitales (Digital Services Act, DSA) representa el intento más avanzado de regular estas dinámicas. Siendo así, puede llegar a

configurarse como un referente normativo frente a posibles implicaciones para el contexto latinoamericano. Este instrumento, que entró en aplicación plena en febrero de 2024, establece un conjunto de obligaciones específicas para los prestadores de servicios intermediarios que tienen particular relevancia para el fenómeno del silenciamiento algorítmico.

Entre las disposiciones más relevantes del DSA para el objeto de este artículo se encuentran: (i) la obligación de las plataformas muy grandes (con más de 45 millones de usuarios activos mensuales en la UE) de realizar evaluaciones anuales de riesgos sistémicos, incluyendo los relativos a los efectos negativos sobre los derechos fundamentales (artículo 33); (ii) la obligación de implementar medidas razonables, proporcionadas y efectivas para mitigar dichos riesgos (artículo 35); (iii) el derecho de los usuarios a no ser sujetos de decisiones de moderación basadas exclusivamente en perfilado automatizado (artículo 17); y (iv) el acceso de los investigadores académicos a los datos de las plataformas para el estudio de los riesgos sistémicos (artículo 40). Este último punto resulta especialmente relevante dado que la opacidad de los algoritmos constituye uno de los principales obstáculos para el estudio empírico del silenciamiento digital.

En este sentido, este Reglamento representa el intento más ambicioso hasta la fecha de regular la actividad de las plataformas digitales en el espacio europeo. De hecho, presta especial relevancia para el presente análisis la disposición 91 del DSA, que permite a las autoridades independientes controlar los sistemas algorítmicos de las grandes plataformas (Parlamento Europeo & Consejo de la Unión Europea, 2022, disposición 91), lo que se aparta de la consideración de que las recomendaciones algorítmicas son "simplemente parte de la infraestructura a través de la cual todo el contenido de sus plataformas es filtrado" (Sáenz Royo, 2025). Esta posibilidad permitiría que los Estados miembros, para garantizar el control efectivo de los sistemas algorítmicos, prevean que las autoridades competentes puedan sancionar los casos en los que se pruebe que un algoritmo privilegia determinados tipos de discurso o discrimina a otros (Parlamento Europeo & Consejo de la Unión Europea, 2022, disposición 91).

Sin embargo, vale la pena mencionar que el DSA presenta limitaciones significativas desde la perspectiva de la protección de los discursos minoritarios. En primer lugar, las obligaciones de transparencia algorítmica, si bien representan un avance notable, no garantizan por sí mismas la equidad en la distribución de la visibilidad del contenido (Cotino, 2023). En segundo lugar, el mecanismo de "notificación y acción" puede incentivar la retirada excesiva de contenido legítimo, especialmente de aquel que proviene de voces minoritarias cuyas formas expresivas se apartan de los estándares comunicativos dominantes. En tercer lugar, la vaguedad de conceptos como "contenido ilícito" o "riesgo sistémico" deja un amplio margen de discrecionalidad a las plataformas, que pueden instrumentar estas categorías para justificar la restricción de discursos incómodos pero legítimos (Revoredo, 2021).

En síntesis, el silenciamiento algorítmico puede entenderse como una manifestación compleja de discriminación indirecta que, además de afectar el derecho a la igualdad, compromete otras garantías fundamentales como el debido proceso o el derecho a la libertad de expresión. Esta

problemática pone en manifiesto la necesidad de repensar los marcos jurídicos y normativos tradicionales para adaptarlos a las dinámicas del ecosistema digital, incorporando estándares de transparencia, responsabilidad y no discriminación en el diseño y funcionamiento de los sistemas algorítmicos, así como un uso ético de los mismos.

### **5.5 Supresión selectiva de contenido o Shadow banning**

El silenciamiento digital, no se produce solamente a través de la eliminación de forma explícita del contenido; sino mediante una diversa serie de mecanismos técnicos de supresión que funcionan de manera invisible para los internautas (Guzmán, 2021). Siendo así, estos aspectos de moderación que se encuentran encubiertos impiden su detección y cuestionamiento, lo que las convierte en herramientas problemáticas desde una visión jurídica y democrática.

De esta manera, uno de los mecanismos más importantes para este contexto es el denominado shadow banning (o “bloqueo en la sombra”). Como ya se mencionó previamente, este consiste en la restricción del alcance de las publicaciones de un usuario de tal manera que este no es notificado (OBSERVACOM, s.f.). Siendo así, el internauta continúa publicando contenido bajo la percepción de que conserva la visibilidad habitual de audiencia, cuando realmente su alcance ha sido significativamente reducido. Lo anterior, funciona como una especie de censura, pero de forma silenciosa (OBSERVACOM, s.f.). Pues bien, en contraste con la eliminación directa de contenido o la suspensión de cuentas (que son medidas visibles, y para ciertos casos, susceptibles de interponer medidas como la apelación), el shadow banning opera de forma opaca, por tanto, limita la posibilidad de interponer algún mecanismo de defensa.

Siguiendo con lo presentado, diversas investigaciones realizadas evidenciaron que diferentes redes sociales están aplicando este mecanismo de shadow banning. Entre estas se encontraba TikTok, la cual aplicaba esta práctica sobre cuentas de activistas LGBTQ+, reduciendo en gran medida la visibilidad de sus publicaciones sin brindar alguna notificación o justificación (Delmonaco et al., 2024). De manera similar, X (antiguamente conocido como Twitter) ha reconocido la existencia de mecanismos de “visibilidad filtrada” (filtered visibility) que reducen el alcance y difusión, de cuentas o contenidos identificados como potencialmente problemáticos a través de sus sistemas automatizados (Delmonaco et al., 2024).

Así pues, el Informe de Transparencia de YouTube de 2021 reconoció la existencia de un mecanismo denominado “borderline content” que reduce la recomendación de contenido que, sin violar las políticas de la plataforma, es considerado “potencialmente problemático”, una categoría cuya definición operativa permanece difusa (Rogers, 2025). Siendo así, la falta de lucidez en los criterios utilizados para esta clasificación introduce un amplio margen de discrecionalidad e incluso, arbitrariedad, lo que dificulta evaluar sus efectos reales. Distintos doctrinantes, como Romero (2024) han expuesto que estas prácticas pueden afectar de manera desproporcionada a publicaciones producidas por colectivos históricamente marginados (como la

comunidad LGBTQ+, mujeres de raza negra, etc) , reforzando dinámicas de exclusión en el entorno digital (Willcox, 2025).

Aunando a lo expuesto, estudios recientes han aportado evidencia cuantitativa sobre estas dinámicas. La investigación de Conti et al. (2024), basado en el análisis de más de 13 millones de publicaciones en el contexto del conflicto que existe entre Ucrania y Rusia, identificó variaciones significativas en la visibilidad del contenido atribuibles a intervenciones algorítmicas. A través de una métrica denominada *p-score*, los autores demostraron que usuarios con niveles similares de popularidad podían experimentar diferencias sustanciales en el alcance de sus publicaciones, incluso en órdenes de magnitud (Conti et al., 2024). Si bien no se encontró evidencia concluyente de discriminación sistemática por orientación ideológica a nivel de comunidad, sí se evidenció que ciertos tipos de contenido (como aquellos que incluían enlaces externos) eran penalizados de manera generalizada, independientemente de su calidad o veracidad (Conti et al., 2024).

De hecho, Nicholas-Williams, Martin y Cameron redactaron una carta a Instagram en 2024, frente a la política de moderación que se emplea. Esto, pues bajo su percepción, es una forma de “contraatacar” las prácticas discriminatorias de eliminación de contenido. De esta manera, argumentaron que: “en un escenario, eliminar imágenes de mujeres negras de talla grande, pero no las de mujeres blancas delgadas, demuestra cómo los sesgos patriarcales y racistas pueden incorporarse a los algoritmos de moderación” (Bonini y Treré, 2024 ). Lo anterior, a pesar de que la red social considera que los algoritmos son imparciales y objetivos.

En virtud de lo expuesto, estos hallazgos permiten determinar que el *shadow banning* son mecanismos efectivos para realizar un control del discurso y que actúan sin la necesidad de recurrir a formas convencionales de censura. De hecho, su carácter de opacidad impide, por un lado, que puedan ser fácilmente detectables; por el otro, que los internautas no puedan comprender o ser conscientes de la disminución en la difusión de sus publicaciones, limitando así su derecho a la igualdad.

Frente a este último punto, desde el punto de vista jurídico, estas acciones pueden generar graves tensiones con derechos fundamentales en caso de que se minimice la visibilidad de contenidos de grupos minoritarios que no infrinjan las políticas claras de agresión. Siendo así, en primer lugar, pueden configurarse un tipo de censura, pero, de manera indirecta; al restringir la circulación de contenidos sin una justificación objetiva y transparente. Esto en concordancia con definiciones de expertos, ya que este tipo de silenciamiento se produce en “un entorno que restringe o desincentiva la libre circulación de opiniones y contenidos.” (Fundación para la Libertad de Prensa, 2025).

En segundo lugar, el *shadow banning* puede producir efectos de discriminación indirecta, en la medida en que afectan de manera desproporcionada a determinados grupos o tipos de discurso (Salomé, 2017). Finalmente, la ausencia de notificación y de mecanismos efectivos de apelación plantea una afectación al derecho del debido proceso el cual en Colombia se encuentra protegido en el artículo 29 de la Constitución Política. Lo anterior, al obstaculizar que los usuarios de las

plataformas digitales puedan ejercer su derecho de defensa frente a decisiones que impactan directamente su participación en el espacio público digital (Sentencia T-256, 2025. C.C.).

En síntesis, la omisión selectiva e injustificada de contenido a través de los algoritmos de las plataformas digitales, puede configurarse como una de las maneras más sofisticadas de silenciamiento a grupos históricamente discriminados. Lo anterior, pues se combina camuflaje, automatización y capacidad de impacto masivo. En este sentido, es imperativo la necesidad de avanzar y poner en discusión, marcos regulatorios que garanticen protección a los ciudadanos que hacen uso de estas herramientas. Adicionalmente, se debe buscar la manera de exigir transparencia y control de las decisiones algorítmicas, con el propósito de proteger diversos derechos fundamentales en el ecosistema digital.

### **5.6 Responsabilidad de las plataformas como “cuasi” reguladores del discurso**

El silenciamiento digital obliga a estudiar la responsabilidad jurídica de las plataformas tecnológicas en su rol de facto como reguladoras del discurso público. Es cierto que estas empresas no se limitan a facilitar la comunicación entre sus usuarios, ya que, además, establecen las condiciones bajo las cuales estas deben ser producidas, circuladas y los requisitos para que adquieran visibilidad (Redem, s.f.). En concordancia con lo expuesto, Tarleton Gillespie (2018) propuso el concepto de “custodios de internet” (custodians of the internet) para designar el poder normativo que ejercen las plataformas sobre las condiciones de la comunicación pública digital. Este se materializa tanto en los términos de servicio y las políticas de contenido como, de manera más determinante, en los sistemas algorítmicos que operativizan dichas reglas en la práctica (Gillespie, 2018).

Siendo así, este escenario es posible analizarlo desde la perspectiva del derecho constitucional comparado. Lo anterior, pues ya existen discusiones sobre una posible extensión de obligaciones propias de los derechos fundamentales a actores privados. Pues bien, el tratamiento jurídico de estas entidades se ha estructurado en torno a la doctrina de la responsabilidad de intermediarios en internet, la cual distingue entre el creador del contenido y quien facilita su difusión (CMS Rodríguez-Azuero, 2022). Bajo este modelo, plataformas, buscadores, proveedores de acceso (ISP) y servicios de alojamiento (*hosting*) no son responsables por los contenidos generados por terceros, salvo que tengan conocimiento efectivo de su ilicitud y, pese a ello, no actúen para retirarlos o bloquearlos (CMS Rodríguez-Azuero, 2022). Este esquema, conocido como régimen de “puerto seguro” (*safe harbor*), ha sido fundamental para el desarrollo del ecosistema digital, al evitar la imposición de cargas excesivas que puedan afectar la libre circulación de información.

Un elemento central de este modelo es el mecanismo de *notice and takedown*, mediante el cual la responsabilidad del intermediario surge sólo cuando, tras ser notificado de un contenido ilegal (por ejemplo, en materia de derechos de autor o difamación), pasa por alto adoptar medidas razonables para su eliminación (Behm, 2024). Este sistema refleja un equilibrio entre la protección de la libertad de expresión y la necesidad de controlar contenidos ilícitos. Al mismo

tiempo, evita imponer a los intermediarios una obligación general de monitoreo previo, lo que podría derivar en formas de censura anticipada.

No obstante, la evolución de las plataformas digitales ha tensionado este esquema clásico. A diferencia de los intermediarios pasivos, las plataformas actuales desempeñan un rol activo en la selección, priorización y difusión de contenidos a través de sistemas algorítmicos (CMS Rodríguez-Azuero, 2022). Esta intervención estructural en la visibilidad del discurso cuestiona la premisa de neutralidad sobre la cual se construyó el régimen de *safe harbor*, y ha llevado a replantear el alcance de su responsabilidad (Behm, 2024).

En el derecho comparado, esta discusión se refleja en distintos modelos regulatorios. En Estados Unidos, la Sección 230 del *Communications Decency Act* otorga una amplia inmunidad a las plataformas frente a contenidos de terceros, incluso cuando realizan ciertas actividades de moderación (Behm, 2024). En Europa, por su parte, el enfoque ha sido más restrictivo, estableciendo que la responsabilidad surge a partir del conocimiento efectivo del contenido ilícito y reforzando progresivamente las obligaciones de diligencia (Behm, 2024). En América Latina, si bien predomina la adopción de esquemas de responsabilidad limitada, se observa una tendencia hacia la incorporación de obligaciones adicionales relacionadas con comercio electrónico, protección de datos y derechos de autor (Behm, 2024).

Ahora bien, este marco se articula con desarrollos del derecho constitucional comparado. En particular, en el ámbito europeo, el Tribunal Europeo de Derechos Humanos, en el caso *Appleby y otros c. Reino Unido* (2003), reconoció que actores privados que controlan espacios relevantes para la expresión pueden, en determinadas circunstancias, estar obligados a garantizar el ejercicio efectivo de derechos fundamentales.

Asimismo, los Principios Rectores sobre las Empresas y los Derechos Humanos (Ruggie, 2011) consolidan esta evolución al establecer que las empresas tienen la responsabilidad de respetar los derechos humanos, lo que implica la adopción de medidas de debida diligencia, evaluaciones de impacto y mecanismos de reparación. Según la OHCHR (2021) expone “El desarrollo y la implantación sistemática de metodologías destinadas a hacer que los sistemas de IA sean más explicables —lo que a menudo se denomina transparencia algorítmica— es de suma importancia para garantizar una protección adecuada de los derechos”. Es decir, esta lógica ha sido extendida al ámbito de la inteligencia artificial, reconociendo que los sistemas algorítmicos (incluidos los de moderación de contenido) deben diseñarse y operarse conforme a estándares de derechos humanos.

Pues bien, la responsabilidad de intermediarios se ve obligada a evolucionar hacia un modelo más complejo, basado no solo en la reacción frente a contenidos ilícitos, sino además en la gestión de riesgos derivados del propio funcionamiento de las plataformas (Arroyo, 2020). Esto resulta, de cierto modo, significativo frente al silenciamiento algorítmico. Pues, según lo que se ha venido abordando a lo largo del artículo, las afectaciones a derechos fundamentales no

provienen necesariamente de contenidos ilegales, sino de decisiones estructurales injustificadas sobre visibilidad, priorización y distribución del discurso.

Sin embargo, este tránsito hacia modelos de responsabilidad reforzada plantea riesgos importantes. La imposición de obligaciones más estrictas puede conllevar a prácticas de sobre-restricción del contenido, trayendo como consecuencia que las plataformas prefieran limitar preventivamente discursos legítimos para evitar posibles sanciones (NetLab UFRJ, 2025). De igual manera, la opacidad de los sistemas algorítmicos dificulta la verificación del cumplimiento de las obligaciones que les recaen, lo que limita la efectividad de los mecanismos de control.

En síntesis, reconocer a las plataformas como “cuasi” reguladores del discurso implica reformular el régimen de responsabilidad de intermediarios. Pues se ve la necesidad de incorporar estándares más robustos de transparencia, debida diligencia y rendición de cuentas, sin desincentivar la libre circulación de información. Siendo así, este equilibrio resulta vital para evitar que el entorno digital se pueda convertir, por un lado, en un espacio de regulación privada sin controles, o por otro, en un lugar de censura extrema proveniente de cargas regulatorias desproporcionadas.

### **5.7. El marco constitucional colombiano y la brecha algorítmica**

En Colombia se configura un gran escenario para analizar el fenómeno del silenciamiento algorítmico, en la medida en que existen diversas condiciones estructurales. En primer lugar, el país cuenta con una de las constituciones más garantistas de América Latina en materia de protección de derechos fundamentales (Ávila, 2017). Esto contrasta de manera evidente con las prácticas de invisibilización en las plataformas digitales, generando una brecha significativa entre los estándares constitucionales y las dinámicas reales de circulación del discurso en ecosistemas digitales.

En segundo lugar, Colombia se caracteriza por tener una diversidad étnica, cultural y lingüística (Marca País Colombia, 2025). Lo anterior, puede llegar a reforzar los efectos de los sesgos algorítmicos, sobre todo aquellos derivados de modelos creados bajo parámetros anglocéntricos (Barret, 2023). Siendo así, esta pluralidad, lejos de ser plenamente reflejada en las plataformas tecnológicas, suele verse marginada dentro de las lógicas de visibilidad predominantes. En tercer lugar, en el país predomina una prolongada historia de violencia, lo que ha ocasionado en muchas ocasiones la censura de voces disidentes (Mejía, 2026). Esto constituye una dificultad en la participación del debate público para estos grupos.

Pues bien, frente a la protección de datos personales, la ley estatutaria 1581 de 2012, constituye el marco normativo fundamental para la garantía de estos. Esta norma establece distintos principios (legalidad, finalidad, libertad, veracidad, transparencia, acceso y circulación restringida, seguridad y confidencialidad). Estos últimos, condicionan el tratamiento de la

información por parte de cualquier entidad que opere en el territorio nacional, incluyendo las plataformas digitales transnacionales. Teniendo en cuenta lo anterior, en el contexto del silenciamiento digital, el derecho al habeas data, el cual se encuentra consagrado en el artículo 15 de la Constitución, adquiere una dimensión que aún el legislador no ha explorado plenamente: el derecho del ciudadano a conocer y cuestionar las decisiones automatizadas que afectan el tratamiento de sus datos y, por ende, su visibilidad en el espacio digital.

Es menester señalar que la Corte Constitucional se ha pronunciado frente a la transparencia algorítmica. En la sentencia T-067, 2025 menciona que esta busca:

*“ (...) que el público en general pueda comprender cómo los sistemas de toma de decisiones automatizadas (SDA) procesan los datos que capturan y cómo toman decisiones que afectan la vida de las personas. Se trata de un principio con un fin constitucional: democratizar el funcionamiento interno de un sistema de toma de decisión automatizado, para que sea entendible por quienes se ven afectados por su puesta en marcha y operación”* (Sentencia T-067, 2025, C.C).

Este planteamiento resulta especialmente relevante en el contexto del silenciamiento digital, en la medida en que pone de relieve la necesidad de hacer visibles los procesos automatizados que inciden en la circulación del discurso. Aunado a lo anterior, la Superintendencia de Industria y Comercio (SIC), como autoridad encargada de velar por el cumplimiento de la protección de los datos personales, ha empezado a explorar la dimensión del silenciamiento digital en investigaciones recientes. Sin embargo, su doctrina expone que el tratamiento algorítmico de datos aún está en proceso de consolidación (Universidad Externado de Colombia, 2024). Lo anterior, evidencia la existencia de ausencias normativas y conceptuales frente a los desafíos que plantean estos sistemas.

En esta línea, la Circular Externa 002 de 2024 de la SIC y el Ministerio de Tecnologías de la Información y Comunicaciones de Colombia (Min TIC), introdujeron lineamientos sobre Inteligencia Artificial y el Tratamiento de datos personales que, a pesar de que no abordan de forma expresa el silenciamiento digital, si establece principios relevantes como: i) el principio de explicabilidad, que exige que las decisiones automatizadas sean explicables de manera comprensible para los afectados; ii) el principio de no discriminación, que prohíbe el uso de datos para producir discriminación arbitraria; y iii) el principio de supervisión humana, que exige que las decisiones automatizadas con impactos significativos estén sujetas a revisión por parte de personas humanas. (MinTIC, 2026). La aplicación consistente de estos principios a los sistemas de moderación de contenidos y de distribución de visibilidad de las plataformas digitales podría constituir un mecanismo eficaz de combate al silenciamiento algorítmico en el marco jurídico vigente.

Ahora bien, frente a las comunidades étnicas colombianas, se presenta un silenciamiento algorítmico e interseccional. En primer lugar, frente a los pueblos indígenas, Colombia reconoce la existencia de 115 pueblos y 65 lenguas nativas (Ministerio de Educación Nacional, 2025), en donde constitucionalmente se encuentran amparados en el marco de la protección de la

diversidad étnica y cultural consagrada en el artículo 7 de la Constitución. Siguiendo con lo expuesto, la Ley 1381 de 2010 conocida como la “Ley de Lenguas Nativas”, establece el deber del Estado de proteger, preservar y fortalecer las lenguas de los grupos étnicos de Colombia. Así mismo, reconoce el derecho de sus hablantes a usarlas en todos los ámbitos de la vida pública (Ley 1381 de 2010, art. 5). Sin embargo, este marco normativo de protección de las lenguas nativas no ha sido extendido al ámbito digital (UNESCO, 2023), creando una brecha que los sistemas algorítmicos de las plataformas llenan con sus propias lógicas que, como se ha documentado, son sistemáticamente desfavorables para las lenguas minorizadas.

Esta brecha normativa tiene consecuencias prácticas directamente relacionadas con el silenciamiento algorítmico. El informe del Center for Democracy and Technology (CDT, 2025), “Content Moderation in the Global South”, que analizó el funcionamiento de los sistemas de moderación de Meta, YouTube e Instagram en lenguas de Latinoamérica, África y Asia, concluyó que los sistemas actuales de moderación basados en procesamiento de lenguaje natural rinden de manera drásticamente inferior en lenguas indígenas sudamericanas (entre ellas el quechua, el aymara y el guaraní) que en inglés o español estándar. El informe señala que el dominio histórico y colonial de la producción de conocimiento en inglés y otras lenguas latinas ha producido datos de entrenamiento que consolidan esa asimetría en los sistemas de inteligencia artificial (CDT, 2025, p. 8). Aunque el informe no analiza lenguas colombianas específicas, sus conclusiones son directamente extrapolables a las lenguas nativas colombianas, dada la comparabilidad de sus características computacionales: muy baja presencia en cuerpos digitales, escasos recursos de procesamiento de lenguaje natural, y ausencia de equipos de moderación especializados en las plataformas.

La UNESCO y Global Voices, en su kit de herramientas “Digital Initiatives for Indigenous Languages” (2023), co-diseñado con redes de activistas digitales indígenas de América Latina, África y Asia, documentan que la polarización de internet en la última década ha producido una reducción progresiva de la disponibilidad y visibilidad de lenguas indígenas en línea. El kit señala que uno de los principales obstáculos que enfrentan los activistas digitales de lenguas indígenas latinoamericanas es la clasificación errónea de sus contenidos por parte de los sistemas de moderación, que frecuentemente los tratan como contenido no reconocido o los someten a revisión bajo categorías de riesgo diseñadas para otras realidades lingüísticas y culturales (UNESCO, 2023, p. 33).

Finalmente, el Banco Interamericano de Desarrollo (BID, 2022) ha identificado además una “triple barrera digital” que afecta específicamente a los pueblos indígenas latinoamericanos: “brecha de energía e infraestructura, brecha de acceso a dispositivos, y brecha de pertinencia cultural” (BID, 2022). Siendo esta última la que se relaciona de forma directa con el problema del silenciamiento algorítmico, ya que incluso cuando las comunidades indígenas logran conectarse a las plataformas digitales, se encuentran con la barrera de que sus lenguas y prácticas culturales no son reconocidas por los sistemas algorítmicos que gestionan su experiencia comunicacional.

## VI. CONCLUSIÓN

De acuerdo con el desarrollo del presente artículo, se demostró la inexistencia y nula neutralidad y objetividad algorítmica en las plataformas tecnológicas de esta nueva era. Con ello, no solo se convierte en un mecanismo de control de masas que avanza de una forma incontrolable e imprescindible, sino que a su vez opera como mecanismos pensados y estructurados para manipular el discurso de todos los grupos y comunidades que hacen parte de una sociedad. Lo más preocupante es que, de manera directa se oculta, invisibiliza, silencia y desvía voces de grupos marciales y minoritarios, cuya necesidad de escucha y visibilidad se torna cada día más fundamental. Con ayuda de un análisis de carácter cuantitativo, a través del cual se hizo el minucioso estudio de artículos, opiniones, jurisprudencia y doctrina, es posible confirmar que la industria tecnológica, a través del tiempo, se ha encargado de reproducir esa ideología discriminatoria en las diferentes plataformas.

### 6.1 Actualidad

En atención a lo planteado anteriormente, respecto de las múltiples ocasiones en las que se ha cuestionado el funcionamiento del algoritmo, es evidenciable que este se encuentra lejos de ser un mecanismo que se desenvuelva de manera aleatoria, y mucho menos neutral. Tanto el “digital Mathew effect” como la teoría de Zuboff que se exploraron a lo largo de este artículo, son un excelente referente para construir y ayudar a describir cómo es que se desenvuelve este fenómeno algorítmico en estos escenarios tan complejos como lo son las redes sociales. Por un lado, confirma efectivamente la manera en la que los grupos sobrerrepresentados y consolidados en la sociedad siempre resultaron beneficiados, mientras que las comunidades marginadas solo se encontraran en un estado de desventaja ininterrumpida, más aún con el auge de herramientas tecnológicas que no les favorece en lo absoluto.

Por otro lado, en concordancia con lo anterior, indiscutiblemente el algoritmo ahora ejerce ese rol de control de masas y discursos. Junto con la minimización del discurso de las minorías y el control de la sociedad, esta es una herramienta que no solo decide ahora qué tipo de contenido circula, que es lo que la población debe ver y que no, sino que a través de eso imponen ideologías, creencias, comportamientos en la sociedad, algunos de ellos, por supuesto, enfocados en mantener minimizado al marginado, pero relevante al sobrevalorado.

A partir de estas conclusiones, realmente lo que termina sucediendo es que se encuentran en conflicto múltiples derechos fundamentales. Entre esos, el que se considera en más conflicto, resulta ser la libertad de expresión. Así, a pesar de que, por ejemplo, la jurisprudencia y normativa colombiana han hecho grandes esfuerzos por resguardar y preservar este derecho en el marco de las redes sociales, sobre todo con pronunciamientos hitos como el de la sentencia T-256 de 2025, se alcanza un punto en el que torna imposible cubrir todos los frentes que el escenario de las redes sociales impulsa. Uno de ellos, sin duda alguna, la regulación del algoritmo en cuanto al silenciamiento de discursos minoritarios, que, en un intento de ser observados y escuchados, este siempre propender a ocultarlo, aun cuando la situación que reflejen sea digna de atención de los jueces de la república y demás espectros sociales. Así, a

pesar de que se haga el intento de incentivar un espacio digital seguro, moderado, basado en criterios de objetividad y transparencia, materialmente hablando no es así.

## **6.2 Iniciativas en el marco regulatorio y normativo**

Si bien, tanto el marco regulatorio nacional como el internacional, a través del cual, por ejemplo, el Pacto Internacional de Derechos Civiles y Políticos, la Constitución Española, La Constitución Política de Colombia, la Convención Americana de Derechos Humanos, entre otros, procuran salvaguardar los derechos fundamentales en entornos digitales, también es cierto que mecanismos como el shadow banning, a través de los cuales se han demostrado la creación de sesgos algoritmos y censura indiscriminada de contenidos, no se ajustan a lo descrito en estos mecanismos normativos. Por el contrario, su enfoque radica en la discriminación directa de múltiples grupos específicos de la sociedad.

Por su parte, en materia de legislación colombiana, con la creación de la Comisión de Regulación de Comunicaciones hizo un intento por regular y controlar ciertas materias en escenarios tecnológicos. Sin embargo, esta no previó el auge y el rápido avance, casi que inminente, de las redes sociales en especial, razón por la cual se estanca en intentar hacerle justicia a este fenómeno. Ahora bien, el Congreso ha hecho múltiples intentos de regular un aspecto fundamental en la época digital, el uso de la Inteligencia Artificial. Así, en sus proyectos de ley 200 de 2023 o 043 de 2025, buscan regular el uso responsable de estas herramientas. No obstante, pierde el foco de lo que le compete a este artículo, el algoritmo, y se evidencia quizás que es un tema que se ha dejado de lado.

En ciertos, aunque pocos, espacios la jurisprudencia busca hacer más relevante el fenómeno del algoritmo, como por ejemplo en la sentencia T-067, 2025 mencionada anteriormente, pero no intenta hacer más profundo. Finalmente, esto termina por colisionar con derechos como el ya indicado, la libertad de expresión, pero también con el manejo de datos personales y demás. Por tal razón, es fundamental que, en materia de derechos, legislación, marco normativo y regulación, las entidades sigan estudiando de forma juiciosa estos espacios, manteniendo la importancia de la circulación de contenido responsable e igualitario.

## **6.3 La necesidad**

Se convierte en un aspecto de especial importancia el hecho de empezar a explorar y conocer a ciencia cierta cómo está trabajando el algoritmo. El funcionamiento tradicional, no regulado ni atendido de este fenómeno produce nada más que irresponsabilidad de los entes competentes en cuanto a la visibilización de todo tipo de circunstancias, discursos, hechos y acontecimientos sociales. La Ley ha sido ineficiente en este sentido, y por el contrario, busca dirigir su foco en ámbitos que parecen ser aún más peligrosos, como la Inteligencia Artificial.

Es menester que como sociedad exista la capacidad de entender el funcionamiento original de los algoritmos, para que en conciencia se pueda originar un espacio de visibilización común, que no discrimine a ciertos grupos y eleven a otros. que el rol de la igualdad se puede observar en todos los escenarios, con el propósito de crear una convivencia más empática.

## VII. BIBLIOGRAFÍA

- Appleby and Others v. the United Kingdom, no. 44306/98, European Court of Human Rights (6 de mayo de 2003). <https://hudoc.echr.coe.int/eng?i=001-61080>
- Arroyo Amayuelas, E. (2020). *La responsabilidad de los intermediarios en internet: ¿Puertos seguros a prueba de futuro?* <https://doi.org/10.20318/cdt.2020.5225>
- Ávila, J. T. (2017). *La teoría del garantismo: poder y constitución en el Estado contemporáneo*. *Revista de Derecho*, (47), 138–166. [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0121-86972017000100138](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-86972017000100138)
- Banco Interamericano de Desarrollo. (2022, agosto 5). *La triple barrera para reducir brechas digitales para pueblos indígenas*. <https://www.iadb.org/es/blog/genero-y-diversidad/la-triple-barrera-para-reducir-brechas-digitales-para-pueblos-indigenas>
- Barton, G., Lee, N. T., & Resnick, P. (2019, mayo 22). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. *Brookings*. <https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/#:~:text=If%20the%20data%20used%20to%20train%20the%20algorithm%20are%20more%20representative%20of%20some%20groups%20of%20people%20than%20others%2C%20the%20predictions%20from%20the%20model%20may%20also%20be%20systematically%20worse%20for%20unrepresented%20or%20under%2Drepresentative%20groups.>
- Barrett, M. (2023, octubre 31). *The dark side of AI: algorithmic bias and global inequality*. Cambridge Judge Business School. <https://www.jbs.cam.ac.uk/2023/the-dark-side-of-ai-algorithmic-bias-and-global-inequality/>
- Behm L., A. (2024). *La responsabilidad de los intermediarios de internet: protección del consumidor y la responsabilidad del intermediario en la comercialización de productos infractores de derechos en el contexto del comercio digital. Un análisis comparativo entre diferentes sistemas jurídicos* (Memoria de pregrado). Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/203926>

- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim Code*. Polity Press.
- Bonini T. Treré E. (2024). *Algorithms of resistance: the everyday fight against platform power*. MIT Press. Available at: [https://owl.purdue.edu/owl/research\\_and\\_citation/apa\\_style/apa\\_formatting\\_and\\_style\\_guide/reference\\_list\\_books.html](https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/reference_list_books.html)
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Bustelo Gracia, J. L. (2024). *Sesgos de género y raciales en la IA: Implicaciones éticas y legales del reconocimiento facial*. <https://doi.org/10.69592/2530-5093-N13-DICIEMBRE-2024-ART-2>
- Calcáneo Monts, M. A. (2021). *Internet, redes sociales y libertad de expresión*. *Cuestiones Constitucionales*, (44), 35–54. <https://www.redalyc.org/journal/885/88571718002/>
- Center for Democracy & Technology. (2025, junio 28). *Content moderation in the Global South: A comparative study of four low-resource languages*. <https://cdt.org/insights/content-moderation-in-the-global-south-a-comparative-study-of-four-low-resource-languages/>
- Children and Young People’s Commissioner Scotland. (s. f.). *What are direct and indirect discrimination?* <https://www.cypcs.org.uk/faq/what-are-direct-and-indirect-discrimination/>
- Civil Liberties Union for Europe. (2021, octubre 7). *Libertad de expresión en las redes sociales: métodos de filtrado, derechos y perspectivas de futuro*. <https://www.liberties.eu/es/stories/libertad-expresion-redes-sociales/43773>
- CMS Rodríguez-Azuero. (2022, abril 25). *Responsabilidad de los intermediarios de Internet*. <https://cms.law/es/col/publication/responsabilidad-de-los-intermediarios-de-internet>
- Comisión Interamericana de Derechos Humanos. (2000). *Declaración de Principios sobre Libertad de Expresión*. <https://www.oas.org/es/cidh/expresion/showarticle.asp?artID=25&lID=2>

- Comisión Interamericana de Derechos Humanos (CIDH) — Relatoría Especial para la Libertad de Expresión. (2013). Libertad de expresión e Internet. OEA/Ser.L/V/II. CIDH/RELE/INF. 11/13.
- Comité de Derechos Humanos de la ONU. (1989). Observación General N.º 18: No discriminación.
- Comunica2, Comunicación y Organización de Eventos. Fontventa. (2021) *Cómo combatir el algoritmo de las redes sociales*.  
<https://comunica2.org/blog/como-combatir-el-algoritmo-de-las-redes-sociales>
- Congreso de la República de Colombia. (2010, 25 de enero). *Ley 1381 de 2010, por la cual se desarrollan los artículos 7º, 8º, 10 y 70 de la Constitución Política, y los artículos 4º, 5º y 28 de la Ley 21 de 1991, y se dictan normas sobre reconocimiento, fomento, protección, uso, preservación y fortalecimiento de las lenguas de los grupos étnicos de Colombia y sobre sus derechos lingüísticos y los de sus hablantes*. Diario Oficial No. 47.603.  
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=38741>
- Congreso de la República de Colombia. (2012, 17 de octubre). *Ley 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales*.  
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>
- Consejo de Derechos Humanos de las Naciones Unidas. (2012). Resolución 20/8: La promoción, protección y disfrute de los derechos humanos en Internet. A/HRC/RES/20/8.
- Constitución Política de Colombia [Const]. 7 de julio de 1991 (Colombia).
- Conti, M., De Cristofaro, E., Galeazzi, A., Paudel, P., & Stringhini, G. (2024). Revealing the secret power: How algorithms can influence content visibility on social media. *arXiv preprint*. <https://arxiv.org/abs/2410.17390>
- Corte Constitucional de Colombia. (2021, 24 de febrero). *Sentencia C-038 de 2021* (M. P. Cristina Pardo Schlesinger, Exp. D-13.752).  
<https://www.corteconstitucional.gov.co/relatoria/2021/c-038-21.htm>
- Corte Constitucional de Colombia. (2024, 25 de octubre). *Sentencia T-453 de 2024* (M. P. Jorge Enrique Ibáñez Najar, Exp. T-9.694.426).  
<https://www.corteconstitucional.gov.co/relatoria/2024/t-453-24.htm>

- Corte Constitucional de Colombia. (2025, 12 de junio). *Sentencia T-256 de 2025* (M. P. Natalia Ángel Cabo, Exp. T-8.764.298).  
<https://www.corteconstitucional.gov.co/relatoria/2025/t-256-25.htm>
- Corte Constitucional de Colombia. (2025, 26 de febrero). *Sentencia T-067 de 2025* (M. P. Natalia Ángel Cabo, Exp. T-8.202.533).  
<https://www.corteconstitucional.gov.co/relatoria/2025/t-067-25.htm>
- Corte Interamericana de Derechos Humanos. (2003). Opinión Consultiva OC-18/03: Condición jurídica y derechos de los migrantes indocumentados.
- Cotino Hueso, L. (2023). *Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida*. *Revista Española de la Transparencia*, (16), 17–63.  
<https://doi.org/10.51915/ret.272>
- Deleuze, G. (1992). Postscript on the societies of control. *October*, Vol. 59.  
[https://cidadeinseguranca.wordpress.com/wp-content/uploads/2012/02/deleuze\\_control.pdf](https://cidadeinseguranca.wordpress.com/wp-content/uploads/2012/02/deleuze_control.pdf)
- Delmonaco, D., Mayworm, S., Thach, H., Guberman, J., Augusta, A., & Haimson, O. L. (2024). “What are you doing, TikTok?”: How marginalized social media users perceive, theorize, and “prove” shadowbanning. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1), Article 154. <https://doi.org/10.1145/3637431>
- Derechos Digitales América Latina. (2025). Resistencia digital en la era de la gobernanza algorítmica: Perspectivas desde la experiencia latinoamericana (N. Solano, Autora; J. Venturini & C. Balla, Eds.).
- Douek, E. (2021). Governing online speech: From “posts-as-trumps” to proportionality and probability. *Columbia Law Review*. Vol 121.  
[https://columbialawreview.org/wp-content/uploads/2021/04/Douek-Governing\\_Online\\_Speech-from\\_Posts\\_As-Trumps\\_To\\_Proportionality\\_And\\_Probability.pdf](https://columbialawreview.org/wp-content/uploads/2021/04/Douek-Governing_Online_Speech-from_Posts_As-Trumps_To_Proportionality_And_Probability.pdf)
- Dvoskin, B. (2023). Expert governance of online speech. *Harvard International Law Journal*. Volume 64.  
[https://journals.law.harvard.edu/ilj/wp-content/uploads/sites/84/HLI102\\_crop-3.pdf](https://journals.law.harvard.edu/ilj/wp-content/uploads/sites/84/HLI102_crop-3.pdf)
- Fundación para la Libertad de Prensa. (2025, agosto 6). *Censura, violencia y estigmatización: amenazas a la prensa libre en Colombia*.

<https://flip.org.co/pronunciamientos/censura-violencia-y-estigmatizacion-amenazas-a-la-prensa-libre-en-colombia>

Foucault, M. (1995). *Discipline and punish: The birth of the prison*. Vintage. (“Panopticism”).

[https://monoskop.org/images/4/43/Foucault\\_Michel\\_Discipline\\_and\\_Punish\\_The\\_Birth\\_of\\_the\\_Prison\\_1977\\_1995.pdf](https://monoskop.org/images/4/43/Foucault_Michel_Discipline_and_Punish_The_Birth_of_the_Prison_1977_1995.pdf)

García-Prieto, V. (s. f.). *Algoritmos y diversidad social: sesgos contra minorías en plataformas digitales*. III Congreso Internacional de Formación Permanente sobre Comunicación Europea (CICOM EU).

<https://cicomeu.org/ponencia/algoritmos-y-diversidad-social-sesgos-contra-minorias-en-plataformas-digitales/>

Gillespie, T. (2014). The relevance of algorithms. En T. Gillespie, P. Boczkowski y K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). MIT Press.

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Guzmán Aguilar, F. (2021, diciembre 9). *¿Cómo funcionan los algoritmos de distribución de contenido en las redes sociales?* Gaceta UNAM.

<https://www.gaceta.unam.mx/como-funcionan-los-algoritmos-de-distribucion-de-contenido-en-las-redes-sociales/>

Habermas, J. (1981). *Theorie des kommunikativen Handelns*. Suhrkamp. [Teoría de la acción comunicativa. Taurus, 1987].

Habermas, J. (1989). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT Press.

IBM. (s. f.). *¿Qué es el sesgo de la IA?* <https://www.ibm.com/es-es/think/topics/ai-bias>

IFEX. (2025, julio 4). *Hablar en clave para existir: resistencias digitales frente a la censura algorítmica*.

<https://ifex.org/es/hablar-en-clave-para-existir-resistencias-digitales-frente-a-la-censura-algoritmica/>

Internet Society Foundation. (2023, septiembre). *¿Qué es la equidad digital?*.

<https://www.isocfoundation.org/es/2023/09/que-es-la-equidad-digital/>

- Jaramillo Gutiérrez, O. J. (s. f.). *Manipulación algorítmica: ¿cómo afecta nuestras decisiones digitales?* Revista Asperger para Asperger. <https://revista.aspergerparaasperger.org/es/manipulacion-algoritmica-como-afecta-nuestras-decisiones-digitales/>
- Korus, K., Salamak, M., & Jasiński, M. (2021). Optimization of geometric parameters of arch bridges using visual programming FEM components and genetic algorithm. *Engineering Structures*, 241(April). <https://doi.org/10.1016/j.engstruct.2021.112465>
- La Rue, F. (2011). Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión.
- Marca País Colombia. (2025). *Colombia, un país que se goza su diversidad étnica*. <https://colombia.co/cultura-colombiana/colombia-pais-de-diversidad-etnica>
- Mejía J., J. G. (2026, febrero 15). *La libertad de prensa bajo violencia: asesinatos, exilios, desplazamientos, agresiones, ataques y amenazas*. Las2Orillas. <https://www.las2orillas.co/la-libertad-de-prensa-bajo-violencia-asesinatos-exilios-desplazamientos-agresiones-ataques-y-amenazas/>
- Ministerio de Educación Nacional. (2025, mayo 29). *Los pueblos indígenas*. <https://www.mineducacion.gov.co/portal/micrositios-institucionales/Sistema-Educativo-Indigena-Propio-SEIP/423418:Los-Pueblos-Indigenas>
- Ministerio de Tecnologías de la Información y las Comunicaciones. (2026, 22 de abril). *Lineamientos de seguridad y privacidad de la información para sistemas de inteligencia artificial* (MSPI). [https://gobiernodigital.mintic.gov.co/seguridadyprivacidad/704/articles-437116\\_recurso\\_1.pdf](https://gobiernodigital.mintic.gov.co/seguridadyprivacidad/704/articles-437116_recurso_1.pdf)
- NetLab UFRJ. (2025, 24 de junio). *Technical note: The “chilling effect” under debate: content moderation on digital platforms post European regulation*. <https://netlab.eco.ufrj.br/en/post/technical-note-the-chilling-effect-under-debate-content-moderation-on-digital>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- OBSERVACOM. (s. f.). *El shadow banning: la censura sutil y oculta de las grandes plataformas digitales. Estudio sobre las prácticas de reducción de alcance de contenidos y cuentas en*

*redes sociales y su relación con la libertad de expresión en línea.*

<https://www.observacom.org/wp-content/uploads/2025/10/Shadowbanning.pdf>

Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (OHCHR).

(2021). The right to privacy in the digital age: Report of the United Nations High Commissioner for Human Rights. A/HRC/48/31.

<https://www.ohchr.org/en/calls-for-input/2021/right-privacy-digital-age-report-2021>

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens

democracy. Crown Publishers. [\(FFLCH\) LIVRO Weapons of Math Destruction - Cathy](#)

[\\_240826\\_220339.pdf](#)

Organización de los Estados Americanos [OEA]. (2013). *Convención Interamericana contra*

*toda forma de Discriminación e Intolerancia*. Adoptada el 5 de junio de 2013 en La Antigua, Guatemala.

Organización de las Naciones Unidas [ONU]. (1966). *Pacto Internacional de Derechos Civiles y*

*Políticos*.

[https://www.ohchr.org/es/instruments-mechanisms/instruments/international-covenant-ci](https://www.ohchr.org/es/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights)

[vil-and-political-rights](https://www.ohchr.org/es/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights)

Organización de las Naciones Unidas, Organización para la Seguridad y la Cooperación en

Europa, Organización de los Estados Americanos, & Comisión Africana de Derechos

Humanos y de los Pueblos. (2011). *Declaración conjunta sobre libertad de expresión e*

*Internet*. <https://www.oas.org/es/cidh/expresion/showarticle.asp?artID=849>

Overbye-Thompson, H., & Rice, R. E. (2025). Understanding how users may work around

algorithmic bias. *AI & Society*, 41(3), 2385-2399.

<https://doi.org/10.1007/s00146-025-02498-1>

Parlamento Europeo y Consejo de la Unión Europea. (2022). Reglamento (UE) 2022/2065

relativo a un mercado único de servicios digitales (Reglamento de servicios digitales).

Diario Oficial de la Unión Europea, L 277, 1–102.

Pasquale, F. (2015). The black box society: The secret algorithms that control money and

information. Harvard University Press.

Revoredo Palacios, A. (2021, febrero 2). *El fenómeno de la censura en redes sociales*. Conexión

ESAN.

<https://www.esan.edu.pe/conexion-esan/el-fenomeno-de-la-censura-en-redes-sociales>

Redem. (s. f.). *¿Cómo funcionan los algoritmos? Así deciden las redes sociales qué contenido enseñarle a tus hijos.*

<https://alfabetizaciondigital.redem.org/como-funcionan-los-algoritmos-asi-deciden-las-redes-sociales-que-contenido-ensenarle-a-tus-hijos/>

Rogers, E. (2025). *The need for greater transparency in the moderation of borderline terrorist and violent extremist content.* *Internet Policy Review*, 14(3).

<https://doi.org/10.14763/2025.3.2012>

Romero Pastor, A. (2024, octubre 30). *The censorship of LGBTQ+ content online corresponds with declines in freedom for everyone.* Tech Policy Press.

<https://www.techpolicy.press/the-censorship-of-lgbtq-content-online-corresponds-with-declines-in-freedom-for-everyone/>

Ruggie, J. G. (2011). Informe del Representante Especial del Secretario General sobre la cuestión de los derechos humanos y las empresas transnacionales y otras empresas comerciales — Principios Rectores sobre las Empresas y los Derechos Humanos. A/HRC/17/31

Sáenz Royo, E. (2025, febrero 26). *¿Está la libertad de expresión en manos de las grandes plataformas?* Agenda Pública.

<https://agendapublica.es/noticia/19670/esta-libertad-expresion-manos-grandes-plataformas>

Salomé Resurrección, L. M. (2017). *La discriminación y algunos de sus calificativos: directa, indirecta, por indiferenciación, interseccional (o múltiple) y estructural.* *Pensamiento Constitucional*, (22), 255–290.

StudySmarter. (s. f.). *Poder simbólico.*

<https://www.studysmarter.es/resumenes/antropologia/antropologia-politica/poder-simbolico/>

Superintendencia de Industria y Comercio. (2024, 21 de agosto). *Circular Externa No. 002 de 2024: Lineamientos sobre el tratamiento de datos personales en sistemas de inteligencia artificial.*

<https://sedeelectronica.sic.gov.co/sites/default/files/normativa/Circular%20Externa%20No.%20002%20del%2021%20de%20agosto%20de%202024.pdf>

- Sustainability Directory. (s. f.). *Algorithmic mediation*.  
<https://lifestyle.sustainability-directory.com/term/algorithmic-mediation/>
- Telefónica. (2021, abril 14). *¿Qué es el mundo digital?*.  
<https://www.telefonica.com/es/sala-comunicacion/blog/mundo-digital-que-es/>
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Tribunal Europeo de Derechos Humanos. (2007). D.H. y otros c. República Checa (GC). Demanda n.º 57325/00. <https://hudoc.echr.coe.int/eng?i=001-83256>
- UNESCO. (2023, octubre 18). *Digital initiatives for Indigenous languages*.  
<https://www.unesco.org/en/articles/digital-initiatives-indigenous-languages>
- UNESCO. (2023, noviembre 16). *Indigenous languages in the digital age*.  
<https://www.unesco.org/en/articles/indigenous-languages-digital-age>
- Universidad Politécnica de Madrid. (2025). *Los algoritmos definen la información que consumimos en redes sociales*.  
[https://www.upm.es/UPM/SalaPrensa/Noticias\\_de\\_investigacion?id=CON17213&fmt=detail&prefmt=articulo](https://www.upm.es/UPM/SalaPrensa/Noticias_de_investigacion?id=CON17213&fmt=detail&prefmt=articulo)
- Universidad Europea. (2025, julio 29). *Engagement: qué es y cómo se calcula*.  
<https://www.universidadeuropea.com/blog/engagement-que-es/>
- Universidad Europea (2025, septiembre 19) *¿Qué es un algoritmo y qué tipos hay? ¿Qué es un Algoritmo? usos y tipos | Blog UE*
- Universidad Externado de Colombia. (2024, agosto 27). *SIC impone nuevas normas: nuevos lineamientos sobre datos personales en inteligencia artificial*.  
<https://www.uexternado.edu.co/proteccion-de-datos/sic-impone-nuevas-normas-nuevos-lineamientos-sobre-datos-personales-en-inteligencia-artificial/>
- Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.
- Wells, A. (2024, abril). *Resistencia, poder, representación y censura algorítmica digitales de las comunidades refugiadas*. Forced Migration Review.  
<https://www.fmreview.org/disrupcion-digital/wells/>

Willcox, M. (2025). *Algorithmic agency and “fighting back” against discriminatory Instagram content moderation: #IWantToSeeNyome*. *Frontiers in Communication*, 9, 1385869.  
<https://doi.org/10.3389/fcomm.2024.1385869>

Zuboff, S. (2019). *The age of surveillance capitalism*. The Fight For a Human Future At The New Frontier of Power. PublicAffairs.  
[https://raggeduniversity.co.uk/wp-content/uploads/2024/08/1\\_x\\_Shoshana-Zuboff-The-Age-of-Surveillance-Capitalism\\_-The-Fight-for-a-Human-Future-at-the-New-Frontier-of-Power-PublicAffairs-Books-2019.pdf](https://raggeduniversity.co.uk/wp-content/uploads/2024/08/1_x_Shoshana-Zuboff-The-Age-of-Surveillance-Capitalism_-The-Fight-for-a-Human-Future-at-the-New-Frontier-of-Power-PublicAffairs-Books-2019.pdf)