ENSAYO

Resistencia digital en la era de la gobernanza algorítmica

Derechos Digitales
AMÉRICA LATINA

perspectivas desde la experiencia latinoamericana

Septiembre, 2025

Resistencia digital en la era de la gobernanza algorítmica
Perspectivas desde la experiencia latinoamericana

Derechos Digitales
AMÉRICA LATINA

Esta es una publicación de Derechos Digitales, una organización independiente sin fines de lucro, fundada en 2005, que tiene como misión la defensa, promoción y desarrollo de los derechos humanos en entornos digitales en América Latina.

Supervisión: Jamila Venturini, Catalina Balla
Investigación y redacción: Nicole Solano
Revisión: Jamila Venturini, Catalina Balla
Diseño: Francisca Balbontín, Nicole Solano

Septiembre, 2025

Introducción

En algún momento, las redes sociales se abrieron paso con una gran promesa para las organizaciones no gubernamentales, movimientos sociales y personas defensoras de derechos humanos como herramientas poderosas para la visibilidad, la movilización y la construcción de narrativas alternativas. A pesar de los diversos contextos sociales y políticos de la región, crearon nuevas oportunidades para denunciar injusticias, convocar a la acción colectiva, compartir conocimientos y amplificar perspectivas tradicionalmente excluidas de los medios de comunicación convencionales. En un contexto en el que el espacio democrático se reducía y la participación pública se restringía cada vez más, estas plataformas se convirtieron en una de las pocas vías que quedaban para ejercer la voz pública.

Ante los intentos de los gobiernos de vigilar y silenciar la expresión en línea (a los que la sociedad civil latinoamericana sigue resistiendo valientemente), hoy en día esa promesa se ve erosionada por otra amenaza creciente: los sistemas automatizados de toma de decisiones que determinan qué contenido es visible, cuáles se ocultan y qué se elimina por completo. Esta moderación algorítmica no solo afecta a quienes publican, sino que también condiciona lo que las audiencias pueden ver, conocer o discutir. Así, lo que alguna vez fue una oportunidad para amplificar voces marginadas, hoy se convierte en un filtro que jerarquiza, distorsiona o elimina discursos no hegemónicos.

Navegar en la ambigüedad de los algoritmos

La libertad de expresión y la protección de las comunidades en línea no deberían ser fuerzas opuestas. Pero la moderación automatizada de los contenidos convierte esa tensión en un silenciamiento o en una modificación forzada de nuestro lenguaje en el entorno digital. La moderación algorítmica silencia las voces críticas y desplaza la memoria de quienes quieren hacerse escuchar, obligándonos a hablar en clave para que nuestras luchas no desaparezcan.

La gestión algorítmica no solo limita el presente, también interviene en la posibilidad de construir futuros más justos. Al condicionar qué narrativas son visibles y cuáles no, los algoritmos afectan directamente la capacidad de articular ideas, reflexiones, desacuerdos y alternativas. Y este impacto se vuelve aún más grave cuando se observa cómo las voces que denuncian violencias estructurales son etiquetadas como “inapropiadas” o simplemente eliminadas. Los discursos que denuncian la injusticia sistémica, la desigualdad o la violencia estatal suelen ser censurados con la excusa de aplicar normas comunitarias opacas y ambiguas.

Las voces que nombran lo que duele, violencias estructurales, desigualdades, crímenes de Estado, son a menudo silenciadas bajo el pretexto de normas comunitarias. De hecho, las personas defensoras de derechos humanos, activistas y periodistas son especialmente afectadas, 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. Esta jerarquización automática de los discursos revela una visión del mundo profundamente sesgada.

La Fundación Friedrich Ebert, una organización alemana con décadas de experiencia en la promoción de la democracia y los derechos humanos

en todo el mundo, incluso a través de asociaciones con actores sociales en América Latina, experimentó esta dinámica de primera mano. Durante años, la Fundación utilizó Instagram para compartir información sobre programas sociales y procesos de selección en toda América Latina. Un día, la plataforma bloqueó la cuenta por supuestas violaciones de sus normas comunitarias, sin ofrecer ninguna explicación específica. Aunque la cuenta se restableció posteriormente, fue suspendida de forma permanente en marzo de 2024, sin previo aviso ni oportunidad significativa de apelar. "Tuvimos que empezar de cero. Perdimos todo nuestro alcance y seguidores", explicó el equipo.

La supuesta infracción se refería a las normas comunitarias sobre «integridad de la cuenta», una categoría ambigua que deja un margen considerable para la interpretación. "Ni siquiera se nos notificó", añadieron. "Solo nos dimos cuenta cuando el sistema dejó de reconocer nuestro correo electrónico". La fundación opera ahora con extrema cautela, plenamente consciente de que incluso un pequeño error en la formulación de su contenido podría dar lugar a nuevas sanciones. Han pedido mecanismos de alerta temprana que permitan a las organizaciones modificar las publicaciones antes de que se apliquen las sanciones, así como una mayor transparencia sobre qué normas se aplican y por qué.

"La aplicación automatizada de las políticas de moderación es inconsistente cuando se trata de casos análogos y carece de comprensión contextual, lo que conduce al silenciamiento y la autocensura de las voces, en particular aquellas que pretenden denunciar la violencia sistémica y estructural mediante un lenguaje y contenidos incisivos, incómodos o denunciatorios", señala Paloma Lara-Castro, directora de Políticas Públicas de Derechos Digitales, quien añade que "esto es especialmente pronunciado en los países de la Mayoría Global, donde se intensifica aún más en contextos sociales críticos como las protestas".

Su análisis destaca cómo las decisiones algorítmicas no solo no aplican estándares equitativos, sino que reproducen y refuerzan sesgos estructurales que afectan de manera desproporcionada a quienes ejercen su derecho a expresarse desde posiciones de resistencia o denuncia. Al hacerlo, contribuyen a la reducción del espacio cívico en línea en regiones que ya se enfrentan a una marginación sistémica.

Las políticas de moderación de contenidos de Meta han demostrado desde hace tiempo un patrón preocupante y constante de supresión de las voces Palestinas, al tiempo que permiten que los contenidos dañinos e incendiarios dirigidos contra los palestinos permanezcan en línea. Esta práctica ha sido especialmente visible en momentos de conflicto agudo. En 2018, 7amleh, el Centro Árabe para el Avance de las Redes Sociales, documentó las prácticas de moderación injustas de la empresa en un informe destacado y, en 2020, lanzó la campaña «Facebook, We Need to Talk» (Facebook, tenemos que hablar). La situación se agravó en 2021, cuando se informó ampliamente de un aumento de las violaciones de los derechos digitales contra el contenido palestino. Se eliminaron masivamente publicaciones, se suspendieron cuentas y se prohibió el acceso a medios de comunicación enteros, silenciando sistemáticamente las narrativas palestinas. Una auditoría de Business for Social Responsibility (BSR) encargada por Meta en 2022 concluyó que las políticas de la empresa habían «afectado negativamente a los derechos humanos de los usuarios palestinos», violando la libertad de expresión, la participación política y la no discriminación.

Esta doble vara, que castiga la denuncia de irregularidades mientras permite los abusos, no es un error técnico. Es el resultado directo de sistemas construidos sin comprensión contextual, marcos de derechos humanos o rendición de cuentas. Se censuran las pruebas, pero la violencia sigue teniendo lugar. Se penalizan las denuncias, pero no las estructuras que las provocan. Al proteger ciertas narrativas y silenciar otras, los algoritmos mantienen el statu quo y obstaculizan las posibilidades de cambio social.

Silenciamiento algorítmico y censura de género: ¿qué voces se están borrando?

Las experiencias compartidas por organizaciones, periodistas y activistas revelan que las normas de las plataformas no se aplican de manera uniforme. En enero de 2024, Instagram desactivó la cuenta de una creadora no binaria que había pasado más de siete años creando y difundiendo contenido sobre educación sexual, reducción de daños y derechos de la comunidad LGBTQ+. La cuenta fue desactivada sin una explicación clara, pero coincidió con la publicación de un contenido sobre Palestina. Según Meta, la suspensión se debió a violaciones relacionadas con «drogas y armas», a pesar de que no se había publicado ningún contenido de ese tipo durante años. "Fue claramente porque hablé de Palestina", escribió al ponerse en contacto con Derechos Digitales para pedir ayuda para restaurar la cuenta, describiendo un progresivo shadowban de su contenido y un fuerte descenso de la visibilidad desde octubre de 2023.

El patrón es revelador: la expresión política en apoyo de determinadas causas suele ser señalada como peligrosa por los sistemas automatizados o las prácticas de moderación ocultas. Como explica la propietaria de la cuenta de Instagram, "las redes sociales se han convertido en un espacio de lucha y resistencia, pero también de injusticia e impunidad"». Su testimonio, compartido por otras personas que han sido censuradas por hablar de Palestina, demuestra cómo la libertad de expresión digital está cada vez más condicionada por los intereses geopolíticos y las narrativas dominantes.

En América Latina, esta asimetría también tiene un impacto significativo en el periodismo feminista. Luciana Peker, periodista feminista argentina, lo sintetiza así: "Las mujeres estamos en la primera línea digital y por eso siempre nos ha tocado poner el cuerpo". En un entorno digital cada vez más hostil, las periodistas que cubren temas de género,

derechos reproductivos o feminismo se enfrentan no solo a abusos coordinados en línea, sino también a sistemas algorítmicos que amplifican el odio y restringen la visibilidad de su trabajo. Los datos son alarmantes: el 80 % de las mujeres periodistas que han sufrido violencia en línea afirman haber limitado su participación en Internet, y una de cada tres ha cambiado de trabajo o ha sido despedida como consecuencia de ello. No se trata solo de un problema individual, sino estructural. La autocensura se convierte en una estrategia de supervivencia en un entorno en el que la exposición puede acarrear daños personales y profesionales.

Peker denuncia además que las plataformas digitales, al no tener protocolos para proteger a las mujeres, están colapsando el periodismo como servicio público. Y no se trata solo de feminismo: "Hoy ya no es sólo un problema de género: es un ataque a la libertad de expresión", sostiene.

La Red Internacional de Periodistas recoge que la moderación de contenidos es un juego asimétrico. En ese laberinto de claves y restricciones, lo que está en riesgo no es solo el alcance inmediato; las denuncias se vuelven más difíciles de rastrear, y la información concreta de organizaciones, activistas o movimientos sociales desaparece para quienes más la necesitan. Como advierte la revista Forced Migration Review, cuando comunidades en condiciones vulnerables se reapropian de términos o usan lenguajes de denuncia, los algoritmos parecen ser más efectivos para eliminar sus contenidos que para moderar discursos de odio o acoso que les atacan. Este desbalance se vuelve crítico cuando miramos quiénes quedan sistemáticamente fuera de las reglas del juego. Las experiencias de personas refugiadas y/o racializadas, por ejemplo, se vuelven aún más invisibilizadas.

El fallo de los algoritmos a la hora de reconocer el contexto de las comunidades del Sur Global agrava la exclusión digital y exacerba las desigualdades en el acceso al discurso público, lo que pone de manifiesto la urgente necesidad de una esfera pública digital que refleje la pluralidad de voces.

La investigación del Center for Democracy & Technology (CDT) sobre la moderación de contenidos en quechua ilustra cómo estos fallos se amplifican en las lenguas no inglesas e indígenas de América Latina. A pesar de que el quechua es uno de los idiomas indígenas más hablados en Sudamérica, el informe reveló que las herramientas de moderación automatizadas, en particular los modelos de lenguaje grandes (LLM), no están entrenadas para comprenderlo, lo que conduce a eliminaciones injustas, una aplicación inconsistente y la propagación del abuso. El estudio destaca que estos retos tienen un marcado carácter de género: las mujeres que se identifican como quechuas en Internet son objeto de acoso de forma desproporcionada, mientras que los equipos de moderación carecen de los conocimientos lingüísticos y culturales necesarios para intervenir de forma eficaz.

Los derechos de autor como herramienta de conveniencia y censura

Las normas de propiedad intelectual se han utilizado durante mucho tiempo para silenciar contenidos incómodos, especialmente cuando denuncian abusos de poder o violaciones de los derechos humanos, y esto también sucede para América Latina. Esta táctica es evidente en el caso de Ponte Jornalismo, un medio independiente brasileño que investiga la violencia policial y el racismo estructural. En 2023, publicaron informes que revelaban que los instructores de una academia de policía enseñaban técnicas de tortura y ejecución. [CB1] Los vídeos que documentaban estas prácticas fueron retirados de YouTube tras las reclamaciones de derechos de autor presentadas por la academia. Irónicamente, sus vídeos originales, que glorificaban estas prácticas, permanecieron en línea.

Ponte apeló, pero ahora corre el riesgo de que se suspenda todo su canal si se presentan nuevas reclamaciones. La Asociación Brasileña de Periodismo de Investigación (Abraji) ha expresado su preocupación por el peligroso precedente que esto sienta: cuando se da prioridad sistemática a la propiedad intelectual sobre el interés público y el acceso a la información, la libertad de prensa se ve gravemente comprometida.

También en Brasil, Intervozes, un colectivo que defiende el derecho a la comunicación, denunció la eliminación abusiva de un vídeo que denunciaba la representación errónea de las mujeres en los programas de televisión por infringir los derechos de autor. No se tuvo en cuenta el uso legítimo a la hora de eliminar producciones que promovían la reflexión crítica sobre los derechos humanos. Se multiplican los casos similares que afectan al activismo con eliminaciones automáticas que ignoran las normas locales e internacionales sobre derechos de autor.

La situación es tan alarmante que las fuerzas del orden han aprendido incluso a utilizar las protecciones de los derechos de autor como arma. Un ejemplo llamativo se documentó en California, Estados Unidos, donde, según se informa, la policía utilizó música protegida por derechos de autor para bloquear la difusión de vídeos que documentaban abusos. Este caso pone de relieve cómo la aplicación de los derechos de autor puede convertirse en una herramienta para obstaculizar el derecho del público a la verdad y la justicia. Por el contrario, los contenidos que promueven ideas misóginas o racistas suelen permanecer en línea sin sanción alguna.

En manos de los poderosos, las normas de las plataformas para hacer cumplir los derechos de autor se convierten en ley, y sus herramientas se utilizan para borrar pruebas, silenciar la disidencia y reforzar el silencio. Cada vez es más evidente que las políticas de moderación de las plataformas distan mucho de ser neutrales. En ausencia de protecciones claras para los periodistas y la sociedad civil, la aplicación de los derechos de autor puede convertirse en una excusa conveniente para suprimir contenidos.

Hablar en código: lenguaje y algospeak

Más allá de las sanciones visibles como la eliminación de cuentas o la limitación del alcance, lo que se castiga es el uso mismo de ciertas palabras. Es en las palabras donde se despliega la potencia de las luchas sociales: nombrar las violencias, exigir derechos, construir memoria. Pero ese mismo poder es leído como amenaza por los sistemas automatizados, que se basan en reglas opacas que ignoran la función política, educativa o crítica del lenguaje.

Como señaló el equipo de la Fundación Friedrich Ebert tras la eliminación de su cuenta en Instagram: “Lo más grave es no saber qué fue lo que se incumplió. No hay manera de anticipar lo que puede activar una sanción.” Esta incertidumbre transforma el lenguaje en una trampa, obligando a organizaciones y activistas a operar con cautela extrema y muchas veces, a autocensurarse antes de publicar.

En respuesta, personas y colectivos han desarrollado estrategias para eludir la censura distorsionando las palabras o utilizando símbolos y eufemismos. En plataformas como TikTok, YouTube o Instagram, es habitual ver expresiones como «4bu$e», «su1c1de» o «desvivir». Esta táctica, conocida como «algospeak», no es solo una estrategia creativa, sino una forma de resistencia dentro de un sistema que penaliza habitualmente el debate sobre verdades incómodas.

El «algospeak» es el resultado de la creatividad y la resistencia en un entorno que redefine lo que se puede decir, impone controles automatizados y reorganiza lo que se considera socialmente relevante. Las consecuencias no son solo individuales: afectan a lo que se puede recordar, a lo que se entiende colectivamente y a cómo las comunidades cuentan sus historias y expresan sus demandas.

Los sistemas automatizados que penalizan las palabras sin contexto obligan a los activistas a reformular su forma de expresarse. Muchos de los términos señalados por los algoritmos han tardado años en definirse y legitimarse dentro de los movimientos sociales. La imposibilidad de nombrar las cosas tal y como son debilita el mensaje y fragmenta la memoria colectiva. Amenaza el núcleo de lo que las democracias latinoamericanas tuvieron que construir para superar una historia de autoritarismo y abuso: el derecho mismo a la verdad y la memoria.

El problema subyacente es que los algoritmos están centralizados en unas pocas corporaciones. Como explicó un usuario de TikTok: "Los sistemas de IA tienen dificultades para comprender la intención detrás de una palabra". Esta limitación técnica se convierte en un mecanismo de silenciamiento estructural, especialmente para temas complejos como la salud, la educación sexual o los derechos humanos. Otro usuario añadió: "Si un vídeo incluye lenguaje sensible, un humano debería evaluar su contexto". Sin embargo, a medida que la moderación se automatiza cada vez más, la intervención humana es poco frecuente, a pesar de los diferentes intentos de hacerla obligatoria.

Aunque la moderación humana es clave, también está lejos de ser perfecta. De hecho, las diferentes demandas judiciales relacionadas con las precarias condiciones laborales de los moderadores de contenido también ponen de relieve la falta de formación adecuada necesaria para aplicar las políticas de la plataforma de manera coherente, lo que a menudo deja el proceso a criterio individual de cada persona que realiza la moderación humana.

Como resultado de la automatización de la moderación y de que esta se lleve a cabo en condiciones precarias, se han producido situaciones como la eliminación de contenido educativo sobre salud sexual, o la penalización de publicaciones que usaban palabras como «sangre». "Prohibir ciertas palabras está creando una sensación de tabú sobre algunos temas", agregó otra usuaria, evidenciando que la censura algorítmica no solo borra contenidos, sino que reconfigura lo que es posible decir.

En este entorno, quienes entienden los códigos pueden evitar la censura. Quienes no lo hacen, quedan excluidos. Digital Future Society advierte que los activistas, los creadores de contenido y los periodistas se ven obligados a reinventar constantemente su lenguaje para seguir siendo visibles en un sistema dominado por la opacidad algorítmica.

Cuando las palabras cambian constantemente para eludir la censura, se vuelve más difícil ponerlas en común y hacerlas accesibles para otras personas. Frente a estas restricciones debemos repensar cómo nos comunicamos y qué estrategias podemos adoptar. Pero esta adaptación tiene un límite: ¿cuánto podemos modificar el lenguaje sin vaciarlo de sentido? ¿Qué pasa con quienes no dominan el código del «algospeak»? ¿Qué voces quedan sistemáticamente excluidas de la conversación digital?

¿Cómo resistimos?

Nombrar con claridad es un acto político. Llamar las cosas por su nombre, sin eufemismos, es también una forma de resistencia. Reapropiarse del lenguaje algorítmico no significa aceptarlo, sino visibilizarlo como síntoma de un sistema injusto. Algunas organizaciones ya crean glosarios y recursos compartidos para descifrar el «algospeak», como el Algospeak Dictionary del Colectivo de Derechos Digitales para la Salud y los Derechos Sexuales y Reproductivos. Democratizar estos conocimientos es clave para que más personas puedan resistir y participar.

También debemos exigir transparencia algorítmica para los países latinoamericanos y de la Mayoría Global, a la par con los estándares ya aplicados en Europa por la Ley de Servicios Digitales (DSA). Las plataformas deben explicar cómo moderan el contenido, bajo qué criterios y con qué sesgos. Esta exigencia no es técnica, es política. Se trata de defender la libertad de expresión y también de proteger las cuestiones políticas, culturales y sociales que están siendo borradas por los sistemas automatizados.

Por último, es urgente construir y apoyar redes que no estén sujetas a los intereses corporativos. Las plataformas descentralizadas como el fediverso, los medios comunitarios, las cooperativas digitales y las alianzas entre organizaciones de la sociedad civil ofrecen alternativas para preservar nuestras historias y mantener la memoria colectiva. La resistencia también significa proteger los espacios donde nuestras palabras no son silenciadas por algoritmos opacos y reclamar nuestro derecho a desarrollar tecnologías alternativas.

Las palabras tienen peso

Las palabras tienen peso. Y cuando las escribimos completas, aunque duelan, aunque molesten, estamos recordando que hay cosas que no pueden ser contenidas, ni siquiera por un algoritmo. Recuperar el lenguaje, cuidarlo y defenderlo es una tarea urgente para quienes luchan por una internet más justa.

En tiempos de automatización y censura silenciosa, reapropiarnos del derecho a nombrar también es reapropiarnos del derecho a existir. No basta con adaptarse: necesitamos transformar el entorno digital para que nuestras palabras no tengan que esconderse. La moderación de contenidos debe ir más allá de los enfoques punitivos y replantearse como una forma de gobernanza participativa, que rinda cuentas a las comunidades a las que afecta y que se centre en la libertad de expresión, la dignidad y la equidad como principios fundamentales, no como consideraciones opcionales.

Más recursos sobre el tema

1. The Santa Clara Principles on Transparency and Accountability in Content Moderation https://santaclaraprinciples.org/
2. Manila Principles on intermediary Liability https://manilaprinciples.org/
3. Disclosure Rules for Algorithmic Content Moderation https://www.hiig.de/wp-content/uploads/2020/12/EoD-Policy-Paper-Blackbox-Breakout.pdf
4. Global Network Initiative https://globalnetworkinitiative.org/

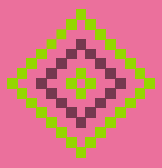
5.Discurso de Odio en América Latina https://www.derechosdigitales.org/wp-content/uploads/discurso-de-odio-latam.pdf

6.Gobernanza algorítmica https://revistalatam.digital/article/22tr07/?pdf=3739

7. Algorithmic content moderation: Technical and political challenges in the automation of platform governance https://journals.sagepub.com/doi/full/10.1177/2053951719897945
8. Lost in Translation: Large Language Models in Non-English Content Analysis https://cdt.org/wp-content/uploads/2023/05/non-en-content-analysis-primer-051223-1203.pdf

9.La moderacíon automatizada en las redes sociales digitales: las movilizaciones lgbt contra la ley Avia en Francia https://journals.openedition.org/ctd/6049

10. The Risk of Racial Bias in Hate Speech Detection https://aclanthology.org/P19-1163/



derechosdigitales.org

Derechos Digitales
AMÉRICA LATINA