

En la mira del algoritmo

Moderación de contenidos
y censura en plataformas digitales



“En el sentido más general, un algoritmo es una serie de instrucciones que indican a una computadora cómo transformar un conjunto de hechos sobre el mundo en información útil. Los hechos son datos, y la información útil es conocimiento para las personas, instrucciones para las máquinas o datos para otro algoritmo”.

Tomado del artículo **¿Qué es un algoritmo?
¿Cómo saben las computadoras qué hacer con los datos?**
The conversation, Jory Denny, octubre 2020.



1
0
01
1
0
1
1
0
01
1
0
1
1
0
01
1
0
1



01
1
0
1
1
0
01
1
0
1
1
0
01
1
0
1

1
0
01
1
0
1
1
0
01
1
0
1
1
0
01
1
0
1

LA ERA DE LA AUDIENCIA ACTIVA

1

Las plataformas digitales se han convertido gradualmente en las nuevas plazas públicas de nuestra sociedad, espacios en los que los ciudadanos pueden discutir y debatir sobre los temas sociales y políticos con un mayor nivel de libertad que en los medios tradicionales como la televisión, la radio o los medios impresos. La creciente adopción de Internet trajo consigo un cambio radical en la forma en cómo se crea y circula la información, haciendo que los usuarios dejen de ser simples consumidores para convertirse en prosumidores (consumidores y productores de información al mismo tiempo).

Sin embargo, tampoco estamos hablando de una utopía, ya que no toda la información que se encuentra en Internet puede ser considerada como cierta. Esta democratización de la forma en cómo se crea la información también abrió nuevas oportunidades para que actores malintencionados utilicen estas herramientas para manipular a las personas según sus propios intereses. La desinformación y la proliferación de noticias falsas han generado la necesidad de que existan grupos de personas y organizaciones dedicadas a la verificación de parte de la información que circula en plataformas digitales y a las empresas detrás de las mismas a tomar acciones para reducir el impacto negativo que puede derivar de esto, algunas veces con mayor éxito que otras.

La remoción de contenido se origina principalmente **a través de 3 vías:**



1. Mediante **solicitudes de remoción de contenidos a las plataformas digitales,** realizadas por actores privados o del Estado.



2. Mediante **comunicación directa con la persona u organización que publicó el contenido** para exigir la eliminación del mismo en sus sitios web o redes sociales.



3. Mediante **la ejecución de los términos de servicio o normas comunitarias de las plataformas digitales.**

Fuente: ARTICLE 19.

La organización ARTICLE 19 define la remoción de contenidos como una práctica para eliminar o restringir la circulación de información en Internet, valiéndose de marcos jurídicos y mecanismos privados que limitan su acceso y explica que esta práctica se origina principalmente a través de tres vías: 1) Mediante solicitudes de remoción de contenidos a las plataformas digitales, realizadas por actores privados o del Estado; 2) Mediante comunicación directa con la persona y organización que publicó el contenido para exigir la eliminación del mismo en sus sitios web o redes sociales; 3) Mediante la ejecución de los términos de servicio o normas comunitarias de las plataformas digitales.

Por supuesto, esta no es una tarea sencilla, según la novena edición de la investigación “Data Never Sleeps”, cada minuto del día se publican en Internet aproximadamente 575.000 tuits, 65.000 fotos en Instagram y 240.000 fotos en Facebook. Moderar este contenido conlleva una serie de desafíos tanto operativos como éticos, ya que se trata de una gran cantidad de información generada en tiempo real, que requiere de una enorme capacidad operativa por parte de las plataformas, que además no existe consenso sobre qué contenido se debe regular o quién debería encargarse de establecer estos criterios y que a su vez puede tener implicaciones negativas sobre derechos fundamentales como la libertad de expresión.

Cada **minuto del día**
se producen en Internet:

575.000 -- tweets

240.000 -- fotos en Facebook

65.000 -- fotos en Instagram



Fuente: Domo “Data Never Sleeps 9.0”

¿POR QUÉ SE MODERAN LOS CONTENIDOS EN LAS PLATAFORMAS DIGITALES? 2

La libertad de expresión no es un derecho absoluto y tiene sus propios límites que previenen que entre en conflicto con otros derechos fundamentales, por lo que algunos contenidos y discursos no están protegidos por el derecho a la libertad de expresión tanto dentro como fuera de las plataformas digitales. Algunos ejemplos de esto son: la propaganda a favor de la guerra, la apología al odio que incite a la violencia, la incitación al genocidio o la pornografía infantil. Esto queda claramente establecido en documentos como el “Marco jurídico interamericano sobre el derecho a la libertad de expresión”, elaborado por la Relatoría para la Libertad de Expresión de la Comisión Interamericana de Derechos Humanos (CIDH).

A su vez, los reguladores de algunos países presionan a las plataformas digitales para que retiren contenido relacionado a bienes, servicios o actos ilícitos, como el terrorismo, la incitación al odio, estafas y fraudes comerciales o violaciones de los derechos a la propiedad intelectual, entre otros. Esto representa un gran desafío si se toma en cuenta que estas plataformas operan a nivel global, por lo que deben respetar y hacer cumplir el marco jurídico de aquellos países en los que mantengan operaciones y que, en ocasiones, pueden ser muy distintos entre sí.

En otros casos la presión también proviene desde la sociedad civil, a través de organizaciones no gubernamentales (ONG) o institutos de investigación que se encargan de documentar prácticas de manipulación en redes sociales y denunciarlas por incumplir con las normas comunitarias de las plataformas. Aunque la decisión final queda en manos de estas últimas, hay ejemplos de éxito que incluso involucran a países como Venezuela; una muestra de esto es que el 2 de diciembre de 2021 Twitter anunció la eliminación de 277 cuentas que amplificaban etiquetas y temas a favor de las narrativas oficiales del Estado venezolano, muchas de estas cuentas habían autorizado una aplicación denominada “Twitter Patria”, la cual también fue suspendida de la plataforma.

La pandemia de COVID-19 también ha obligado a las plataformas a tomar un rol más activo en el control de la infodemia y la lucha contra la desinformación en redes sociales, que de no ser atendidas pueden llegar a representar un peligro para la salud pública. Esto se puede observar en las campañas junto a organismos multilaterales como la Organización de Naciones Unidas, pero también en la remoción de contenidos que incluso ha afectado publicaciones realizadas por jefes de Estado como Jair Bolsonaro o Nicolás Maduro.

EL JUEGO DE LAS PLATAFORMAS

3

No todas las razones para moderar o censurar ciertos tipos de contenido son jurídicas o provienen de organismos externos a las plataformas, algunas veces estas obedecen a decisiones corporativas que ocasionan que se censure contenido lícito como los desnudos y el contenido sexual. En algunos casos, como el de Facebook e Instagram, se fundamenta esta decisión bajo el argumento de que este tipo de contenido puede resultar sensible para algunos usuarios, aunque expertos opinan que también se trata de una motivación económica para no perder anunciantes. En otros casos, como el de la plataforma Only Fans, un intento de retirar el contenido sexualmente explícito de la misma provino de la presión de aliados comerciales clave como socios bancarios y proveedores de plataformas de pago.

Otras plataformas como Telegram han dado señales mixtas en cuanto a su intención de moderar los contenidos a los cuales pueden acceder los usuarios a través de sus aplicaciones. Por un lado, la empresa anunció a principios de 2021 que había bloqueado “docenas de canales públicos con llamadas a la violencia para miles de suscriptores”, hecho que coincidió con la decisión de otras plataformas como Twitter, Facebook y Google de cerrar las cuentas del expresidente de los Estados Unidos, Donald Trump, y combatir los discursos que incitaban a la violencia, sobre todo luego del asalto al Capitolio ocurrido días antes.

Sin embargo, meses después la empresa anunció una versión para Android con “menos restricciones” que puede ser descargada directamente desde su sitio web y que no estaría sujeta a los términos y condiciones de la tienda de aplicaciones de Google. A pesar de que la compañía no dejó claro a qué se refería con “menos restricciones”, en algunos de los canales que han bloqueado se muestra un mensaje que explica que los mismos no pueden ser visualizados en la versión de la tienda de Google, dejando entrever que estos aún podrían ser accesibles a través de la aplicación que se descarga directamente desde su sitio web.

Otro caso interesante es el de la plataforma Parler, una red social que dice diferenciarse de las demás por no censurar el contenido que comparten sus usuarios, lo que la popularizó entre sectores conservadores como los seguidores del expresidente Donald Trump. La carencia de moderación de Parler con respecto al contenido ilícito o que puede ser considerado peligroso le costó el ser eliminada de las tiendas de aplicaciones de Google y Apple e incluso la suspensión del servicio de alojamiento que mantenía con Amazon. El hecho ocurrió días después del asalto al Capitolio de los Estados Unidos y de que algunas investigaciones independientes revelaran el papel que jugó la plataforma en la organización de los involucrados en este hecho, pero también por denuncias que señalaban que esta red social servía como espacio para el racismo y otros tipos de discurso de odio.

Actualmente la plataforma se encuentra nuevamente operativa e incluso su aplicación está disponible para ser descargada desde la tienda de Apple, aunque en sus normas comunitarias (actualizadas al 14 de febrero de 2021) especifican que eliminarán aquellos contenidos que se utilicen como herramienta para cometer delitos, agravios civiles u otros actos ilícitos.

UN DESAFÍO: LA MODERACIÓN DE CONTENIDOS 4

Los tiempos que vivimos plantean una serie de retos para la humanidad en muchos aspectos, incluyendo los legales, con tecnologías que avanzan mucho más rápido que los marcos regulatorios que se deben aplicar a las mismas. De hecho, ni siquiera entendemos completamente cuáles son las implicaciones para los derechos humanos de muchas de las tecnologías que llevamos años utilizando.

Recientemente Meta (empresa matriz de Facebook, Instagram y WhatsApp) anunció que desactivará el sistema de reconocimiento facial en Facebook y eliminaría las plantillas con datos biométricos de más de mil millones de personas por considerar que su uso debe ser evaluado “considerando preocupaciones crecientes sobre el uso de esta tecnología en general”. Pero estas preocupaciones están lejos de ser nuevas, en 2016 expertos ya alertaban que los algoritmos pueden replicar y ayudar a perpetuar los prejuicios y la discriminación que ya existe en nuestra sociedad, aunque no sería hasta 2020 que esta discusión llegaría a hacerse masiva con el estreno del documental Coded Bias en Netflix.

Pero nuestra vida ya está llena de algoritmos que utilizamos prácticamente para todo lo que hacemos en Internet. Durante el inicio de la pan-

demia por COVID-19 las grandes plataformas de redes sociales anunciaron que delegarían una mayor parte de sus tareas de moderación de contenido en algoritmos, lo que inmediatamente se tradujo en un aumento de los errores en cuanto a publicaciones indebidamente marcadas o retiradas de estas plataformas, pero también de la suspensión de cuentas en Twitter, que en el caso de Venezuela afectó a organizaciones no gubernamentales como el Observatorio Venezolano de Prisiones y la Comisión para los Derechos Humanos del Estado Zulia, así como también a medios de comunicación como el canal de televisión Vale TV y el diario Tal Cual, entre otros.

LA PENALIZACIÓN SELECTIVA

5

Los problemas de la moderación automatizada de contenido no comenzaron con la pandemia, el caso de las protestas de Chile en el año 2019 ya nos daba señales de lo que en los próximos años se convertiría en un problema recurrente. En 2021 esta discusión volvió a estar en el centro de la opinión pública por la eliminación de publicaciones durante el Paro Nacional en Colombia y el desalojo de ciudadanos palestinos del barrio Sheikh Jarrah en Jerusalén; ante esto, Instagram emitió un comunicado desde su cuenta de Twitter en el cual vinculaba ambos incidentes a un error técnico de su plataforma. Además, días antes, coincidiendo con denuncias públicas sobre casos de acoso y abuso sexual en Venezuela realizadas a través de redes sociales, la etiqueta “#YoSiTeCreo” fue temporalmente restringida en Facebook.

Algunos matices en el tipo de contenido que está permitido en las plataformas también pueden llegar a ser muy complicados para los algoritmos. Por ejemplo, a pesar de que el contenido sexualmente explícito está prohibido en la mayoría de las plataformas, con el tiempo se han flexibilizado las normas para permitir ciertas categorías específicas para publicaciones con fines educativos, de salud, artísticos y en el caso de Facebook incluso humorísticos o satíricos. Sin embargo, en la práctica se siguen cometiendo errores que afectan a obras de arte que contienen desnudos, vídeos e imágenes sobre educación sexual e incluso videos que buscan fomentar la prevención del cáncer de mama.

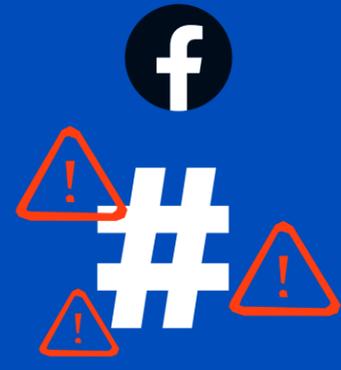
La moderación automatizada de contenidos ha generado un aumento de los errores en:



Suspensión de cuentas de Twitter



Eliminación de publicaciones e historias en Instagram



Restricciones de etiquetas en Facebook

Fuentes: propias, Instagram y Access Now.

En otros casos, los intentos de censura provienen de actores malintencionados que intentan aprovecharse de la moderación automatizada para afectar contenidos legítimos. En Venezuela, el medio digital El Pitazo ha sufrido la suspensión de su canal de YouTube en cuatro ocasiones distintas como consecuencia de haber publicado videos de carácter informativo, dos de las suspensiones estaban relacionadas a alocuciones en las que Nicolás Maduro afirmaba que Venezuela produciría tratamientos contra el COVID-19, aunque sin ningún tipo de respaldo científico, otra provino de un video en el que se mostraba un ataque terrorista en la ciudad de Bagdad y la última de ellas relacionada a un video publicado en 2019 en el que se muestra a una mujer maltratando a menores de edad.

En este último vídeo, el medio afirmó que se evitó mostrar las caras de las víctimas y que la viralización del mismo fue fundamental para que las fuerzas policiales pudieran identificar a la victimaria; esta fue la tercera sanción que El Pitazo recibió en menos de tres meses, razón por la cual YouTube anunció que el cierre del canal sería definitivo. El medio apeló esta decisión alegando que varios medios de comunicación internacionales también habían publicado los mismos videos sin recibir ningún tipo de sanción y luego de un proceso de revisión la decisión fue revertida y el canal restituido.

Anteriormente El Pitazo ha sido víctima de ataques de denegación de servicio distribuidos (DDoS) en dos oportunidades y de múltiples bloqueos a por lo menos cuatro de sus dominios desde 2017 tanto en proveedores de Internet públicos como privados, por lo cual se presume que las denuncias contra el canal de YouTube podrían ser un nuevo intento de censurar el trabajo del medio en Internet.

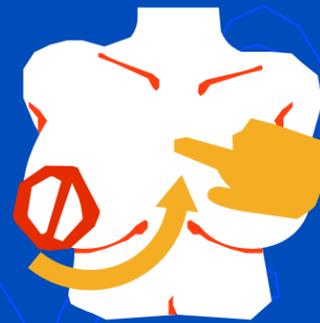
La moderación de contenidos también afecta contenido educativo como:



Obras de arte con desnudos



Material de educación sexual



Contenido sobre cuidado de la salud

Fuentes: La Razón, Euronews, El Mundo.

En otros países como Estados Unidos, algunos miembros de las fuerzas policiales han reproducido música con derechos de autor para evitar que los videos donde los graban puedan ser compartidos en plataformas como Instagram y YouTube. En uno de los casos un policía dijo a un manifestante luego de reproducir una canción “puedes grabar todo lo que quieras, solo sé que no podrás publicarlo en YouTube” y minutos después lo reafirmó a petición del manifestante diciendo “estoy reproduciendo música para que no puedas publicar [el video] en YouTube”.

Pero las redes sociales no son el único espacio en el que se utilizan prácticas cuestionables para tratar de censurar contenido, en algunos casos empresas y particulares utilizan métodos no automatizados y muy elaborados que combinan manipulaciones legales y técnicas para tratar de engañar a medios de comunicación y que estos eliminen de sus sitios web investigaciones o artículos que denuncian casos de corrupción u otros crímenes, así como también a motores de búsqueda como Google para que desindexen los enlaces a los mismos y en algunos casos incluso a proveedores de alojamiento para que suspendan el servicio a estos medios, dejando sus sitios web inaccesibles al menos temporalmente.

¿QUIÉNES SE ENCARGAN DE MODERAR LOS CONTENIDOS?

6

Hace unos años, las relaciones públicas y la gestión de la reputación eran asuntos de preocupación solo para las grandes empresas y la farándula. Con el Internet y las redes sociales todo dio un vuelco, los medios de comunicación tradicionales fueron ampliados, sustituidos y/o fagocitados por el ecosistema digital y en este nuevo escenario las personas expresan constantemente sus opiniones acerca de cualquier tópico y tema y las comparten en línea todo el tiempo.

En esta dinámica las marcas, empresas, instituciones y personas contratan agencias y profesionales que les garantizan la pulcritud de su imagen. Se trata de un servicio personalizado que se denomina gestión de la reputación en línea u ORM por sus siglas en inglés, una especialidad que utiliza diversas estrategias para arreglar, proteger y construir reputación, suprimiendo información negativa, promoviendo información positiva, y valiéndose de una batería de recursos para establecer y generar confianza y credibilidad.



Los primeros casos documentados del uso de estas prácticas en Venezuela se remontan a 2016, cuando la empresa española Eliminalia solicitó la eliminación de artículos publicados en los medios de comunicación El Pitazo, Runrunes, El Estímulo, A Todo Momento y el portal Poderopedia, cuyo objetivo es mapear las relaciones de poder en los negocios y la política venezolana, en ese año las solicitudes se fundamentaban principalmente en interpretaciones del marco jurídico nacional, incluyendo la Constitución venezolana. El portal Poderopedia afirma haber recibido al menos otras cinco solicitudes de remoción de contenido durante los primeros 10 meses del año 2021, tres de ellas realizadas a título personal y otras dos a través de un bufete español cuyo nombre prefirieron no revelar por motivos de seguridad.

En junio de 2020, el medio de investigación Armando.info también recibió una solicitud de eliminación de uno de sus reportajes para que este no apareciera en los resultados de las búsquedas relacionadas al nombre de una de las personas mencionadas en el mismo, justificando la petición

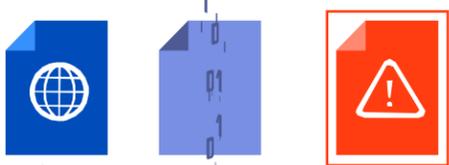
con artículos de la Constitución venezolana combinados con legislaciones europeas. Posteriormente, con el apoyo de la ONG sueca Qurium, el medio pudo confirmar que el correo con la solicitud, enviado desde una dirección que aparentaba pertenecer a alguna autoridad legal europea (abuse@legal-abuse.eu), realmente habría sido enviado por Eliminalia; Armando.info afirmó haber recibido al menos tres correos similares desde el año 2019, todos provenientes de servidores de esta empresa.

REMOCIÓN DE CONTENIDOS: 7 TECNOLOGÍA + EXPERTICIA + ESTRATEGIA

Esta no es la única técnica utilizada por Eliminalia, en abril de 2021 Qurium publicó dos investigaciones en las que muestra cómo esta empresa presentó entre mayo de 2019 y marzo de 2021 al menos 92 reclamos falsos amparados en la Ley de Derechos de Autor de la Era Digital (DMCA, por sus siglas en inglés), promulgada por Estados Unidos y cuyo objetivo es regular y proteger los derechos de autor en Internet, entre estos casos se incluían 16 relacionados con Venezuela. Estos reclamos tenían como objetivo que Google (principal motor de búsqueda en Internet) desindexara los resultados de artículos publicados por medios de comunicación en los cuales se denuncian distintas formas de corrupción y crimen organizado.

En julio de 2021 Armando.info recibió varias notificaciones de Google en las que les informaban sobre la desindexación de tres de sus reportajes por reclamos de derechos de autor sustentados en la DMCA. De igual forma, en octubre de 2020 el medio de comunicación La Gran Aldea informó que habían recibido correos solicitando la eliminación de uno de sus artículos y semanas después una notificación de su proveedor de alojamiento sobre una denuncia por plagio fundamentada en la DMCA; esto se sumaba a una serie de ataques de DDoS de los que su sitio web fue víctima semanas antes de la notificación. Otro caso similar es el de ONG Acceso a la Justicia, que sufrió una suspensión del servicio de alojamiento de su sitio web tras recibir dos denuncias por plagio basadas en la DMCA durante diciembre de 2019 y marzo de 2020.

Metodología utilizada para **desindexar de Google enlaces a artículos e investigaciones:**

- 
1. Copian el contenido del sitio web original y lo publican en uno distinto (generalmente creado solo con este fin).
- 
2. Modifican la fecha de publicación para que parezca anterior a la del artículo original.
- 
3. Presentan un reclamo por supuesta infracción de la DMCA ante Google.

Fuentes: Qurium.

Las sanciones derivadas de estos errores, afectan a medios de comunicación y organizaciones no gubernamentales que a menudo no cuentan con los conocimientos suficientes para comprender la razón por la que están siendo penalizadas y que además en contextos como el venezolano son objetivo de ataques tanto en el ámbito físico como digital. Pero el mayor daño colateral es sobre el derecho de acceso a información de interés público, ya que este tipo de medidas han demostrado ser parcialmente eficaces para quienes desean borrar parte de su pasado de Internet, especialmente la vinculada a actos ilícitos, por lo que es fundamental que las empresas proveedoras de alojamiento y los motores de búsqueda como Google sean más transparentes con sus procesos y revisen exhaustivamente los reclamos por infracción de derechos de autor.

Existen otros casos en los que la censura de las plataformas no es accidental, lo que pone en tela de juicio la tesis de quienes afirman que estas pueden y deben autorregularse. En agosto de 2021 Facebook cerró las cuentas de varios investigadores de la Universidad de Nueva York alegando una violación de sus términos de servicio, con esta decisión la empresa puso fin al trabajo del equipo que estudiaba cómo los usuarios son seleccionados con base en sus intereses para decidir qué anuncios electorales se les muestra dentro de la plataforma. Acciones como estas hacen crecer la desconfianza, sobre todo luego del escándalo de Cambridge Analytica, las recientes denuncias de ex empleados y filtración de documentos de la compañía en los denominados “Facebook Files”.

UN MONO CON TIJERA: REGULACIONES Y CONTROLES

8

Esto a su vez ha hecho que algunos gobiernos se manifiesten a favor de mayores controles hacia las plataformas, una tendencia que creció durante el último año según reveló la edición más reciente del informe Freedom on the Net, ya que al menos 24 países (incluyendo Venezuela) anunciaron o impulsaron la creación de nuevos reglamentos para regular la forma en que estas compañías tratan los contenidos publicado por sus usuarios.

Pero los expertos afirman que:

“Permitir a las autoridades estatales o a los intermediarios regular contenido puede traducirse en el silenciamiento de voces disidentes, en especial en sociedades autoritarias”.

En China, a pesar de ser actualmente el país con el Internet menos libre del mundo, empresas como Apple, Google y Zoom han intentado mantenerse en el mercado dando al gobierno un mayor control sobre los datos de sus usuarios, censurando el contenido que se muestra en sus plataformas e incluso cerrando las cuentas de activistas y defensores de los derechos humanos y la democracia de este país.

Independientemente de que las grandes plataformas cooperen o no, muchos regímenes no democráticos han impulsado leyes y decretos para restringir la libertad de expresión a través de mecanismos técnicos y legales. En el caso de Venezuela, esto ocurrió con la polémica “Ley del Ciberespacio”, rechazada por casi 90 organizaciones nacionales e internacionales, que se intentó promulgar desde la Asamblea Nacional Constituyente (ANC) en el año 2019 y que a principios de 2021 se volvió a mencionar dentro de un paquete de 33 leyes que Nicolás Maduro quiere impulsar desde la nueva Asamblea Nacional, incluyendo también una reforma a la “Ley de Responsabilidad Social en Radio, Televisión y Medios Electrónicos” para incluir un capítulo sobre las redes sociales.

Estos instrumentos vendrían a acompañar la denominada “Ley Constitucional contra el odio, por la convivencia pacífica y la tolerancia”, aprobada en 2017 por la ANC y que, a pesar de “carecer de los requisitos formales y materiales para ser una ley” según la ONG Espacio Público, ha sido utilizada para imponer la censura y autocensura en el país mediante la detención arbitraria y procesamiento judicial de ciudadanos por publicaciones

en plataformas digitales y medios tradicionales, tal como lo demuestran diversas investigaciones.

Un caso que hay que destacar es el del ex presidente Donald Trump, quien en mayo de 2020 firmó una orden ejecutiva para modificar la “Ley de Decencia de las Comunicaciones”, una legislación de Estados Unidos destinada a regular el contenido de Internet. Específicamente, la sección 230 de esta ley protege a los intermediarios al establecer que no serán responsables por el contenido publicado por terceros en sus plataformas, lo que hace posible que el Internet tal y como lo conocemos hoy exista. Aunque esto parecía una medida de presión política ante la remoción de publicaciones del expresidente por parte de Facebook y Twitter, si hubiese sido aprobada esta modificación podría haber dado lugar a más contenido dañino dentro de las plataformas u ocasionado que las mismas tuviesen que filtrar todo el contenido publicado por terceros, con las implicaciones para la libertad de expresión que eso conlleva.

ESTÁNDARES PARA LA MODERACIÓN DE CONTENIDOS EN INTERNET

Los enfoques centrados solo en la regulación o autorregulación de las plataformas en Internet son insuficientes para enfrentar los problemas de la moderación de contenidos, por lo que es necesario que tanto empresas como gobiernos avancen hacia un modelo de corregulación en el que ambos se complementen y permitan adoptar acuerdos sobre mecanismos y estándares claros, involucrando en las discusiones a organizaciones de la sociedad civil y considerando sus recomendaciones durante el proceso de toma de decisiones para garantizar un enfoque multisectorial como el que caracteriza a la gobernanza de Internet.

Adicionalmente, el uso de algoritmos en la moderación de contenidos debe estar sujeto a la supervisión y el control humano, que garantice el funcionamiento adecuado del mismo y pueda intervenir para proteger los derechos de los usuarios en caso de que sea necesario.

A continuación, se enumeran algunos estándares para la adopción de políticas y prácticas sobre responsabilidad de intermediarios que sean respetuosas de los derechos humanos, extraídas de los **Principios de Manila:**

1. Los intermediarios deberían estar protegidos por ley de la responsabilidad por contenido de terceros.

2. No debe requerirse la restricción de contenidos sin una orden emitida por una autoridad judicial.

3. Las solicitudes de restricción de contenidos deben ser claras, inequívocas y respetar el debido proceso.

4. Las leyes, órdenes y prácticas de restricción de contenidos deben cumplir con las pruebas de necesidad y proporcionalidad.

5. Las leyes, políticas y prácticas de restricción de contenido deben respetar el debido proceso.

6. La transparencia y la rendición de cuentas deben ser incluidas dentro de la normativa, políticas y prácticas sobre restricción de contenido.

Para profundizar sobre este tema se pueden consultar documentos elaborados por expertos de diversos sectores como las “26 recomendaciones sobre gobernanza de contenido” de Access Now, los “Estándares para una regulación democrática de las grandes plataformas” de Observacom, el informe “Mirando Al Sur” del consorcio Al Sur, los “Principios regulatorios para la formulación de políticas” de Twitter y el informe “A/HRC/38/35” del Ex-Relator Especial de las Naciones Unidas sobre la libertad de opinión y de expresión, David Kaye.

CRÉDITOS

DIRECCIÓN EJECUTIVA
Marianela Balbi

COORDINACIÓN DE LIBERTADES INFORMATIVAS
Daniela Alvarado Mejias

REDACCIÓN E INVESTIGACIÓN
David Aragort

CONSULTORÍA COMUNICACIONAL
Claudia Machillanda

DISEÑO Y VISUALIZACIÓN DE DATOS
Camila Agelvis

