



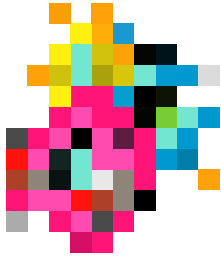
# **26 recomendaciones sobre gobernanza de contenido:**

**una guía para legisladores, reguladores y  
encargados de políticas empresariales**



Access Now (<https://www.accessnow.org>) defiende y extiende los derechos digitales de los usuarios en riesgo alrededor del mundo. Mediante la combinación de apoyo técnico directo, campañas globales, el análisis integral de políticas públicas, el financiamiento a grupos locales emergentes y eventos como RightsCon, luchamos por los derechos humanos en la era digital.

Para obtener más información sobre este informe, póngase en contacto con Javier Pallero ([javier@accessnow.org](mailto:javier@accessnow.org)).

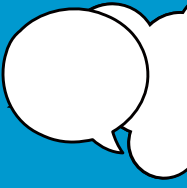


Este documento es una publicación de Access Now escrita por Eliška Pírková y Javier Pallero, con la colaboración del equipo de políticas públicas de Access Now. Los autores agradecen especialmente las contribuciones de Ben Wagner, Daphne Keller, Mathias Vermeulen, Kate Klonick, Aleksandra Kuczerawy, Brett Solomon, Fanny Hidvégi, Guillermo Beltrà, Juliana Castro, Donna Wentworth, y el equipo de políticas públicas de Access Now.

# Contenidos

<b>I. Resumen ejecutivo</b>	<b>7</b>
<b>II. Introducción: cómo ha evolucionado la regulación de Internet</b>	<b>8</b>
LOS INICIOS Y EL SURGIMIENTO DE LA REGULACIÓN	8
DESAFÍOS PARA LOS DERECHOS DE LOS USUARIOS EN LÍNEA: INTENSIFICACIÓN DEL PROBLEMA	8
<b>III. Gobernanza de contenido: cómo se establecen normas y quién las aplica</b>	<b>10</b>
REGULACIÓN ESTATAL	10
AUTORREGULACIÓN	10
MODERACIÓN DE CONTENIDO	10
CURACIÓN DE CONTENIDO	11
CORREGULACIÓN	11
<b>IV: Riesgos: cómo afectan las decisiones de gobernanza de contenido a los derechos humanos</b>	<b>12</b>
RIESGOS DE LA REGULACIÓN ESTATAL PARA LOS DERECHOS HUMANOS	13
RIESGOS DE LA AUTORREGULACIÓN PARA LOS DERECHOS HUMANOS	15
RIESGOS DE LA CORREGULACIÓN PARA LOS DERECHOS HUMANOS	17
<b>V. Guía: 26 recomendaciones para una gobernanza de contenido respetuosa de los derechos humanos</b>	<b>19</b>
<b>A. Recomendaciones para una regulación estatal respetuosa de los derechos humanos</b>	<b>21</b>
1. ATENERSE A ESTRICTOS PRINCIPIOS DEMOCRÁTICOS	21
2. ADOPTAR MECANISMOS DE PUERTO SEGURO Y EXENCIONES DE RESPONSABILIDAD LEGAL	22
3. NO IMPONER UNA OBLIGACIÓN DE MONITOREO GENERAL	24
4. DEFINIR MECANISMOS DE RESPUESTA ADECUADOS	25
5. ESTABLECER NORMAS CLARAS PARA CUANDO NO APLIQUEN LAS EXENCIONES DE RESPONSABILIDAD LEGAL	26

6. EVALUAR DETENIDAMENTE Y DE FORMA LIMITADA EL CONTENIDO DE ILEGALIDAD MANIFIESTA	26
7. CREAR PROCEDIMIENTOS DE NOTIFICACIÓN Y ACCIÓN QUE RESPETEN LOS DERECHOS FUNDAMENTALES	27
8. LIMITAR LAS MEDIDAS TEMPORALES E INCLUIR SALVAGUARDAS	29
9. IMPONER SANCIONES PROPORCIONADAS PARA EL INCUMPLIMIENTO	30
10. USAR MEDIDAS AUTOMATIZADAS SOLO EN CASOS LIMITADOS	31
11. LEGISLAR SALVAGUARDAS PARA EL DEBIDO PROCESO	32
12. CREAR OBLIGACIONES SIGNIFICATIVAS DE TRANSPARENCIA Y RENDICIÓN DE CUENTAS	33
13. GARANTIZAR LOS DERECHOS DE APELACIÓN Y REPARACIÓN EFECTIVA DE LOS USUARIOS	35
<b>B. Recomendaciones para una autorregulación respetuosa de los derechos humanos</b>	<b>35</b>
1. PREVENIR PERJUICIOS A LOS DERECHOS HUMANOS	36
2. EVALUAR EL IMPACTO	36
3. SER TRANSPARENTES	36
4. APLICAR LOS PRINCIPIOS DE NECESIDAD Y PROPORCIONALIDAD	37
5. CONSIDERAR EL CONTEXTO	37
6. NO TOMAR POSICIONES ARBITRARIAS NI DE INJUSTA DISCRIMINACIÓN	38
7. PROMOVER LA TOMA DE DECISIONES CON INTERVENCIÓN HUMANA	38
8. CREAR MECANISMOS DE NOTIFICACIÓN Y REVISIÓN	39
9. BRINDAR REPARACIONES	40
10. INVOLUCRARSE EN LA GOBERNANZA ABIERTA	41
<b>C. Recomendaciones para una correulación respetuosa de los derechos humanos</b>	<b>41</b>
1. ADOPTAR MARCOS LEGALES PARTICIPATIVOS, CLAROS Y TRANSPARENTES	42
2. NO CAMBIAR NI TERGIVERSAR LAS RESPONSABILIDADES DE LOS ACTORES	42
3. PREVENIR ABUSOS	42
<b>VI. Conclusión</b>	<b>43</b>
<b>VII. Glosario</b>	<b>44</b>
<b>Notas finales</b>	<b>48</b>



# I. Resumen ejecutivo

Internet representa una herramienta esencial para ejercer los derechos humanos, incluidos el de acceso a la información y la libertad de opinión y expresión, entre otros. Los servicios que actúan como intermediarios para el flujo de la información, especialmente las plataformas como los servicios de redes sociales y los motores de búsqueda, cumplen también un rol importante. En teoría, las plataformas digitales les dan a todas las personas la oportunidad de conectarse, pero los gobiernos represores pueden obstruir esa capacidad mediante legislación que pone en peligro los derechos de las personas.

Al mismo tiempo, las normas que usan las plataformas para regular el contenido y la actividad de los usuarios —normalmente formuladas de manera unilateral— están, por lo general, diseñadas y aplicadas de maneras que entran en conflicto con la libertad de expresión, la privacidad y otros derechos fundamentales. A su vez, es posible que esto habilite nuevas formas de explotación, tanto por parte de actores privados como públicos.

Las acciones que llevan a cabo en esta área las plataformas y los gobiernos, junto con aquellas que no logran llevar a cabo, pueden perjudicar a las sociedades y, particularmente, a las poblaciones vulnerables. Los fenómenos sociales que son cada vez más comunes, como las campañas de desinformación o contenido ilegal que pueden incitar al odio o a la violencia, se pueden manifestar de varias formas y con diversas características que plantean una amplia gama de desafíos. Su impacto y las consecuencias que tienen sobre los derechos humanos difieren significativamente según el contexto geográfico, cultural y político. En algunos lugares del mundo, las consecuencias negativas y el daño colectivo a grupos afectados pueden representar o causar atrocidades.

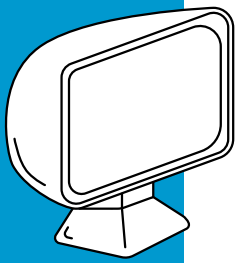
Por lo tanto, este documento no busca establecer un conjunto universal de soluciones específicas para los problemas complejos y delicados que surgen a partir de la gobernanza de contenido.

En cambio, sobre la base de nuestra experiencia en el desarrollo de políticas públicas alrededor del mundo, ofreceremos una guía básica centrada en los derechos humanos que sirve como referencia esencial para políticas de gobernanza adecuadas, dado que las partes interesadas deben tener en cuenta a los actores y las tecnologías específicas de su región. Nos explayaremos más sobre la implementación regional y temática de estas recomendaciones en futuras publicaciones.

Este documento comienza con una definición de los tres principales tipos de estructura de gobernanza que se usan en la actualidad: la regulación estatal, la autorregulación y la corregulación.

Luego, exploraremos los riesgos que cada uno de estos tipos plantea para los derechos humanos y ofreceremos recomendaciones para abordar esos riesgos. Cuando sea posible, recomendaremos políticas de base concretas para abordar problemas clave como la responsabilidad legal de los intermediarios, las medidas automatizadas, las decisiones de autorregulación, etc.

Nuestro objetivo es llegar a los tomadores de decisiones en todo el mundo para que se prioricen los derechos humanos en todos los debates sobre gobernanza de contenido. Esta es la única forma de crear un futuro digital que refuerce los ideales comunes de libertad, apertura, y los valores democráticos, con el potencial de devolverles el poder a los usuarios.



## II. Introducción: cómo ha evolucionado la regulación de Internet

### LOS INICIOS Y EL SURGIMIENTO DE LA REGULACIÓN

Cuando los primeros servicios de redes sociales hicieron su debut mundial a fines de la década de 1990 y principios de los 2000, la opinión popular —de los expertos en tecnología, los analistas y los usuarios de Internet en general— era que Internet iba a habilitar que las personas compartieran y tuvieran acceso a más información que nunca. Para muchos, el potencial del acceso y el intercambio de información ilimitados a nivel mundial indicaba que Internet empoderaría a los usuarios y actuaría como una fuerza que destrabaría la acción colectiva y la democratización.<sup>1</sup> En efecto, muchos individuos y organizaciones a nivel mundial han sido capaces de aprovechar Internet como herramienta para lograr un impacto real, usando el activismo en línea para defender los derechos humanos y fortalecer la democracia. Access Now y otras organizaciones de derechos digitales trabajan para proteger y preservar esta capacidad. Sin embargo, con el desarrollo de la Internet comercial, los gobiernos de todo el mundo comenzaron a reconocer y responder a los problemas que surgen cuando se comparte el contenido en línea, incluido aquel que se considera ilegal.

Muchos de los primeros esfuerzos regulatorios se centraron en las violaciones de derechos de autor y el material pornográfico en Internet.<sup>2</sup> Fueron menos prominentes los esfuerzos para abordar los fenómenos sociales que dominan el debate sobre políticas públicas en torno a la Internet de hoy en día, como el discurso de odio, la difamación o el contenido terrorista. La atención inicial sobre los derechos de autor y la pornografía hizo que las plataformas en línea adoptaran, de manera voluntaria, la primera generación de filtros de carga para la detección de contenido ilegal.

Si bien las primeras huellas de la regulación de Internet tomaron su forma a partir de la preocupación de los gobiernos y la industria sobre los derechos de autor y el material pornográfico, los intermediarios<sup>3</sup> desarrollaron mecanismos de autorregulación para lidiar con los problemas que surgían de otras formas de expresión en línea, como el discurso de odio, el acoso, el hostigamiento y la pornografía no consensuada, entre otras. Hoy en día, seguimos lidiando con las primeras influencias de una gobernanza desarrollada con una mentalidad impulsada por la propiedad intelectual, en contraposición a un enfoque más holístico que considere todo el espectro de problemas causados por el discurso en línea. La bifurcación de los enfoques se ha trasladado a la toma de decisiones, tanto en el sector público como en el privado. El creciente uso de herramientas automatizadas para regular el discurso en línea, desarrolladas para la aplicación de los derechos de autor pero que amenaza la libre expresión, remarcó la fragmentación de los enfoques.

### DESAFÍOS PARA LOS DERECHOS DE LOS USUARIOS EN LÍNEA: INTENSIFICACIÓN DEL PROBLEMA

La manera en que los usuarios acceden e intercambian información ha cambiado considerablemente desde los inicios de Internet. No solo se ha expandido el acceso a Internet, sino que el surgimiento de las plataformas de redes sociales ha simplificado y aumentado la cantidad de interacciones. Esto implica más acceso a la información y una expansión del goce de la libertad de expresión y otros derechos fundamentales.<sup>4</sup>



También ha traído consigo nuevas dificultades, como la creación y diseminación de contenido que los gobiernos consideran ilegal, a una escala sin precedentes. También ha habido un incremento en el contenido que, aunque posiblemente no sea ilegal en algún país en particular, es considerado "dañino" o no deseado, ya sea por parte de las plataformas en línea, sus usuarios, o los reguladores de los Estados nación. El material utilizado para el ciberacoso, la propagación de la desinformación, o el contenido violento entran en esta categoría.<sup>5</sup> En este sentido, las plataformas en línea han comenzado a utilizar la autorregulación, mediante la creación y aplicación de normas sobre la expresión considerada aceptable en sus servicios.

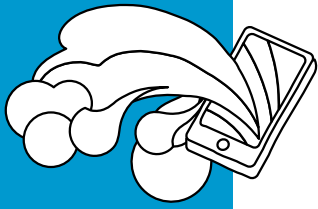
Estas normas internas, comúnmente llamadas "términos de servicio" o "pautas de la comunidad", entre otras denominaciones, normalmente incluyen una lista de comportamientos que están prohibidos en la plataforma. Definen cuáles son los contenidos y comportamientos aceptables y cuáles no. Es posible que también expliquen y justifiquen los principios y valores de la plataforma, como se consagran en las pautas.

Debido a la creciente presión que ejercen los usuarios de plataformas y los reguladores para tomar más medidas contra la propagación de contenido ilegal y dañino generado por los usuarios, las plataformas en línea han reforzado cada vez más sus normas con respecto a todos los tipos de contenido, incluido el discurso de odio y el material implicado en la incitación a la violencia.<sup>6</sup> Sin embargo, el proceso que usan las plataformas para implementar estas normas autoimpuestas ha sido, peligrosamente y en gran medida, opaco. De hecho, entre los actores involucrados en la gobernanza de Internet, los actores privados son quienes revelan la menor cantidad de información sobre cómo se formulan y aplican sus mecanismos de regulación.<sup>7</sup>

Las preocupaciones de los Estados nación sobre la propagación de contenido ilegal en todas las plataformas en línea se justifica con razón. Resulta urgente encontrar respuestas adecuadas a los complejos fenómenos sociales como el discurso de odio o la radicalización en línea. No obstante, esto no puede servir de excusa para imponer una censura excesiva sobre el discurso de los usuarios. Los enfoques y las herramientas usados por gobiernos y plataformas deben respetar los estándares internacionales de derechos humanos y el estado de derecho, asegurando que los derechos y las libertades fundamentales de los usuarios ocupen un lugar central.

Si las soluciones de gobernanza desarrolladas por los gobiernos y las plataformas (mediante regulación y autorregulación, respectivamente) son poco aconsejables, apresuradas, o no incorporan los principios y las salvaguardas internacionales para los derechos humanos, pueden aumentar el riesgo para los usuarios, en lugar de disminuirlo.

Si bien los gobiernos y las plataformas luchan contra las presiones para responder a crisis de relaciones públicas, influencias del mercado o quejas de los grupos de interés, los más vulnerables en esta dinámica son los usuarios. La misión de Access Now es defender y extender los derechos digitales de usuarios en riesgo en todo el mundo, y esto incluye asegurar que los individuos y los grupos en riesgo no se conviertan en víctimas de censura o abuso en línea, ya sea mediante políticas gubernamentales o empresariales. Nuestro equipo trabaja internacionalmente, con oficinas y personal en muchas regiones del mundo, y queremos resaltar la necesidad de que exista una mayor comprensión de las sutiles diferencias de este debate en función del contexto social, cultural y legal. Mediante nuestra Línea de Ayuda de Seguridad Digital<sup>8</sup> y nuestro trabajo en coalición, estamos en contacto directo con activistas, periodistas, defensores de derechos humanos y otras personas afectadas negativamente por el contenido ilegal o no deseado. Estos mismos usuarios están en la mirilla de las regulaciones gubernamentales y las prácticas de autorregulación de las empresas que, aunque tengan buenas intenciones, han dañado su capacidad de libre expresión y empoderamiento. Esperamos que esta guía pueda darles a los tomadores de decisiones los cimientos para reivindicar sus derechos y preservar sus libertades.



### III. Gobernanza de contenido: cómo se establecen las normas y quién las aplica

Los actores de este ecosistema de comunicaciones en línea cada vez más complejo tienen el deber de tener en cuenta los derechos humanos. Los gobiernos están obligados a proteger estos

derechos, y las empresas tienen la responsabilidad de respetarlos. En el contexto de estos deberes y responsabilidades, analizamos cómo se crean las normas gubernamentales y quiénes las hacen cumplir.

En este documento, el término "gobernanza" hace referencia a los complejos procesos e interacciones que crean e involucran a partes interesadas públicas y privadas para hacer cumplir las normas que rigen el contenido en línea.<sup>9</sup> Dividimos la gobernanza de contenido en tres categorías principales: **regulación estatal**, impuesta por los gobiernos; **autorregulación**, ejercida por las plataformas; y **corregulación**, ejercida en conjunto por los gobiernos y las plataformas mediante acuerdos vinculantes o voluntarios.

Cada uno de estos tipos presenta desafíos únicos para la protección de los derechos humanos en línea, como se describe a continuación.

#### REGULACIÓN ESTATAL

La regulación estatal hace referencia a cualquier instrumento vinculante legal o regulatorio que promulgan las instituciones públicas locales, nacionales o regionales mediante sus procesos legislativos o regulatorios. Normalmente, los estados definen el contenido en línea que se considerará ilegal en sus códigos penales. En algunas jurisdicciones, los estados también recurren a la formulación de legislación específica para regular el contenido, como en el caso de los derechos de propiedad intelectual.

#### AUTORREGULACIÓN

Las plataformas en línea definen qué tipo de contenido es aceptable al usar sus servicios, a menudo mediante la creación de sus propios términos de servicio.<sup>10</sup> Estos términos pueden contener definiciones que incluyan el contenido ilegal y el contenido que, a pesar de ser legal, es considerado indeseable.

Este es un ejercicio de autorregulación que se puede ser unilateral o llevarse a cabo luego de la consulta con los usuarios u otras partes interesadas. Principalmente, las plataformas llevan a cabo la autorregulación de dos formas: mediante la moderación y la curación de contenido.

#### MODERACIÓN DE CONTENIDO

La moderación de contenido es una práctica mediante la cual una empresa que posee y administra una plataforma en línea que aloja o brinda acceso a contenido generado por usuarios<sup>11</sup>—como un motor de búsqueda o un servicio de red social— toma decisiones sobre alojar o continuar alojando una porción de contenido específico en virtud de sus términos de servicio.

Una decisión respecto de alojar contenido o no hacerlo puede implicar dar de baja el contenido permanente o temporalmente, ya sea en toda la plataforma o en relación con ciertos grupos de usuarios en un área geográfica específica.

## CURACIÓN DE CONTENIDO

Las decisiones sobre el alcance, la prominencia o la amplificación de cierto contenido (conocidas como "curación de contenido"<sup>12</sup>) determinan cuántos y cuáles grupos de usuarios están expuestos al contenido y la manera en la que este se presenta. Estas decisiones se toman sobre la base de diferentes criterios y mediante distintos métodos.<sup>13</sup> Las plataformas en línea generalmente recomiendan contenido a los usuarios mediante sus páginas de inicio o sugerencias personalizadas. Las decisiones de curación de contenido pueden implicar impulsar el alcance y la exposición de algunas formas de discurso, o minimizar o limitar ese alcance.

## CORREGULACIÓN

La corregulación es un enfoque cada vez más popular que complejiza aún más la gobernanza de contenido. Los regímenes de corregulación se pueden definir como una autorregulación que las autoridades públicas promueven, apoyan y, en ocasiones, monitorean activamente. Algunos ejemplos típicos de los mecanismos de corregulación son los códigos de conducta voluntarios que resultan del diálogo entre actores privados y autoridades nacionales o regionales.

Los mecanismos de corregulación pueden tener distintos niveles de formalidad. Las iniciativas de corregulación, en el sentido estricto, a menudo incluyen un elemento regulatorio formal —como una ley o una resolución administrativa— que actúa como marco legal y rige las actividades de los actores involucrados, incluyendo las normas y las diferentes consecuencias.<sup>14</sup> Sin embargo, en otros casos, la cooperación entre autoridades públicas y actores privados está regida por acuerdos voluntarios informales que también establecen normas y objetivos. Algunos ejemplos son el Llamamiento de ChristChurch para Eliminar el Contenido Terrorista y Extremista Violento de la Red<sup>15</sup> y el Foro Global de Internet contra el Terrorismo,<sup>16</sup> entre otros.

Debido a que estas iniciativas plantean muchas de las mismas preocupaciones sobre derechos humanos, en este documento, usaremos el término "corregulación" en el sentido amplio para referirnos a cualquier iniciativa en la que tanto los gobiernos como los actores privados cumplan un papel, ya sea que estén o no vinculados a marcos regulatorios o legales subyacentes.



## **IV. Riesgos: cómo afectan las decisiones de gobernanza de contenido a los derechos humanos**

La gobernanza de contenido que es incompatible con las normas y principios básicos de derechos humanos pone en peligro la libertad de expresión, el acceso a la información, la libertad de opinión y asociación, la privacidad y otros derechos humanos. También afecta a diferentes poblaciones de distintas maneras. A continuación, describimos los desafíos que diferentes tipos de usuarios enfrentan y los riesgos más prominentes vinculados a mecanismos comunes de gobernanza de contenido.

### **LAS DECISIONES DE GOBERNANZA AFECTAN A DIFERENTES CATEGORÍAS DE USUARIOS DE DISTINTAS MANERAS**

En teoría, las plataformas les dan voz a todas las personas, pero el discurso dañino también puede acelerarse en entornos conectados a la red. Un sinnúmero de personas se vieron afectadas por la amplificación, la proliferación y la ubicuidad del discurso ilegal o dañino. Las víctimas no forman parte de un grupo homogéneo, sino que incluyen a poblaciones marginalizadas, como las mujeres, minorías religiosas o étnicas, personas de color, y miembros de la comunidad LGBTQ.<sup>17</sup> Irónicamente, estos individuos y grupos están generalmente descentralizados y dependen de las plataformas en línea para encontrar a otras personas con los mismos intereses, conseguir apoyo, intercambiar información, expresar sus opiniones, participar en la política, organizarse, trabajar y acceder a la cultura, entre otras formas en las que se ejercen los derechos fundamentales. Perder acceso al contenido que han publicado, ser silenciados o eliminados de espacios de discurso y deliberación públicos puede causar graves impactos negativos en sus vidas y constituye una pérdida terrible de espacio para el debate público.

Lo mismo sucede en el caso de las decisiones de gobernanza de contenido que afectan a organizaciones de la sociedad civil, activistas y periodistas. Al preparar respuestas regulatorias para la gobernanza de contenido, los estados deberían armonizar la protección de los datos personales con el derecho a la libertad de expresión y el acceso a la información, especialmente en relación con el periodismo, la investigación científica, o la expresión artística o literaria.

Para proteger estas actividades, los marcos legales deben contemplar exenciones o derogaciones de prohibiciones generales en lo que respecta a categorías especiales de contenido en línea.

Los tomadores de decisiones de gobernanza de contenido, ya sean parte de gobiernos o de empresas, deben asegurar un cuidado y una atención extra en el diseño y la implementación de las salvaguardas de los derechos humanos de grupos vulnerables. Por consiguiente, las políticas que afecten a dichos grupos deberían garantizar su participación efectiva en las etapas de diseño e implementación.

### **Las decisiones de moderación de contenido son una preocupación cada vez mayor para los usuarios en riesgo**

La Línea de Ayuda de Seguridad Digital de Access Now, que trabaja con personas y organizaciones para mantener su seguridad en línea, ha sido testigo de un aumento en la cantidad de casos relacionados con las decisiones de moderación de contenido que afectan a los usuarios que están más expuestos a riesgos.

Los problemas relacionados con la moderación de contenido representaron aproximadamente el 20% de los casos en la Línea de Ayuda de Seguridad Digital en 2019<sup>18</sup> (~311 casos). Generalmente, estos casos se pueden clasificar en dos categorías: recuperación de contenido y baja de contenidos. Los casos de recuperación de contenido incluyen situaciones en las que los usuarios consideraron que su contenido fue dado de baja injustamente. Esto puede deberse a errores en la aplicación de los términos de servicio de la plataforma, que generalmente derivan de problemas con los sistemas de detección automatizados, pero también de errores cometidos por moderadores humanos. Los casos de baja de contenido incluyen solicitudes de los clientes para asistencia en la baja de contenido que, a menudo, tiene que ver con el acoso, la suplantación de identidad, la publicación maliciosa de información privada, y la incitación a la violencia o amenazas de muerte dirigidas a defensores de los derechos humanos, periodistas, y otros miembros de la sociedad civil como represalia por su trabajo. En muchos de estos casos, los usuarios se ponen en contacto con la Línea de Ayuda de Seguridad Digital de Access Now como medida de último recurso, luego de que las solicitudes que envían a los canales oficiales de las plataformas de redes sociales sean ignoradas o denegadas.

## **RIESGOS DE LA REGULACIÓN ESTATAL PARA LOS DERECHOS HUMANOS**

13

Existen preocupaciones específicas en relación con el ejercicio del poder estatal. El riesgo de que el gobierno cometa abusos siempre es una posibilidad, y no solo en países gobernados por regímenes autoritarios. Por lo tanto, los estándares de derechos humanos, el estado de derecho, las instituciones democráticas efectivas, las políticas basadas en evidencia, la transparencia y la participación son esenciales en toda regulación de la expresión. En algunas propuestas regulatorias y legislación aprobada, estas salvaguardas están ausentes, lo que representa un riesgo concreto para los derechos de los usuarios.

### **PROBLEMAS CON LA LEGALIDAD, LA NECESIDAD Y LA PROPORCIONALIDAD**

Los actores estatales a menudo tienen inquietudes legítimas sobre la diseminación de contenido ilegal en línea. Lamentablemente, los reguladores estatales, por lo general, no cumplen con los principios legales fundamentales a los que están obligados. Las leyes que son incompatibles con los derechos humanos sirven, generalmente, como pretexto para exigir que las plataformas eliminen rápidamente contenido alojado en sus sitios, lo que puede reprimir el discurso y la disidencia legítimos.<sup>19</sup> Algunos estados recurren a leyes penales restrictivas desproporcionadas que contienen definiciones amplias de crímenes como extremismo, difamación, blasfemias, y otros actos relacionados con el discurso. Por ejemplo, a principios de 2019, Australia modificó su código penal en tan solo 48 horas para añadir una sentencia de prisión para los ejecutivos de negocios de plataformas en línea que no dieran de baja el "contenido extremista" dentro de un plazo acotado.<sup>20</sup>

En otros casos, los gobiernos han propuesto y adoptado legislación que se abusa de la autorregulación y las medidas voluntarias empleadas por los actores privados.<sup>21</sup> Este enfoque es incompatible con las salvaguardas básicas de los derechos humanos y, en última instancia, conduce a la remoción excesiva de contenido o directamente a la censura.

Los modelos regulatorios estatales deberían centrarse específicamente en el contenido de explícita ilegalidad y evitar la regulación de definiciones de fenómenos sociales en línea que están en constante cambio, como la desinformación, el discurso de odio y el contenido terrorista.

Las restricciones del derecho a la libertad de expresión deben estar claramente prescritas por ley, perseguir un objetivo legítimo, ser necesarias en una sociedad democrática y ser proporcionadas respecto del objetivo establecido.

El Tribunal Europeo de Derechos Humanos determinó en su jurisprudencia<sup>22</sup> que cualquier expresión, incluido el contenido en línea, debe ser evaluado en el contexto y las circunstancias en las que se disemina, teniendo en cuenta su posible impacto, los medios de diseminación, los autores, y su influencia pública.

Inherentemente, esa prescripción contiene la idea de que las definiciones legales demasiado amplias llevarán, en última instancia, a injerencias no necesarias y desproporcionadas en el derecho a la libertad de expresión.

### **DELEGACIÓN DE FUNCIONES JUDICIALES A LAS PLATAFORMAS**

Existe una tendencia en la que la legislación se centra en delegar algunas funciones típicamente estatales a las plataformas. La Ley de Aplicación de Redes de Alemania, conocida también como NetzDG,<sup>23</sup> impone la obligación de que las plataformas evalúen y eliminen contenido que se considera ilegal en virtud del derecho penal alemán, tarea que normalmente cumple un funcionario independiente.

Los mecanismos de este tipo son peligrosos para los derechos humanos por varias razones. Los objetivos de una empresa privada son muy diferentes de aquellos que tiene un funcionario público y democrático. Un funcionario judicial o administrativo independiente debe actuar dentro de los límites de la legislación nacional correspondiente y las leyes de derechos humanos. Tal funcionario debe considerar cuál es el mejor plan de acción para alcanzar un objetivo legítimo<sup>24</sup> para un estado democrático y tener en cuenta los principios de necesidad y proporcionalidad.

Los principios fundamentales recogidos en las constituciones políticas contemplan la necesidad de la independencia económica y política de las autoridades decisoras. También es necesario que cuenten con conocimiento experto del asunto en su competencia. Finalmente, los sistemas de adjudicación brindan recursos de apelación y reparaciones.

En la práctica, a veces es difícil aplicar estas características y salvaguardas. Es necesario contar con respuestas rápidas, y los gobiernos generalmente tienen dificultades en cuanto a recursos, la creación de capacidad, y otros desafíos prácticos. No obstante, necesitamos reforzar y reformular el acceso a la justicia siguiendo pautas jurídicas basadas en derechos humanos. La mayoría de las características que garantizan que se priorice el interés

público no están presentes en los procesos de toma de decisiones de las plataformas privadas.

### **FALTA DE EVIDENCIA E INFORMACIÓN AL MOMENTO DE REGULAR**

Toda regulación estatal que aborde los fenómenos sociales en línea como la desinformación, el discurso de odio o el contenido terrorista debe estar siempre fundada en evidencia contundente. Muchas propuestas legislativas presentadas o adoptadas recientemente por gobiernos nacionales de todo el mundo buscan combatir el contenido indeseado sin respaldo apropiado ni factual que demuestre el impacto que puede causar el contenido en los grupos vulnerables o la relación entre el comportamiento en línea y fuera de línea. Este es el caso de la legislación que propone dar de baja sitios y servicios completos para luchar contra la desinformación y el odio en línea en Honduras<sup>25</sup> and Bangladés.<sup>26</sup>

El riesgo de regular con la intención de alcanzar un resultado en particular sin comprender completamente los principios básicos que se verían comprometidos es, actualmente, un problema en todos los gobiernos. Sin el conocimiento necesario, la astucia política no da buenos resultados, y provoca aún más daños. La falta de políticas basadas en evidencia condujo a la adopción de respuestas regulatorias apresuradas que se centran en la rápida remoción de contenido y en la cantidad de contenido eliminado. La regulación que apunta a la velocidad y la cantidad de remoción de contenido puede, a su vez, generar un cumplimiento excesivo por parte de las plataformas en línea que resulte en bajas ilegítimas de contenido generado por los usuarios.<sup>27</sup>

### **FALTA DE UN ENFOQUE OPORTUNO Y PARTICIPATIVO**

A menudo, los gobiernos reaccionan a los fenómenos sociales apresurándose para adoptar nueva legislación, sin dar lugar a un debate previo suficiente y sin consultar con las partes interesadas relevantes. Esto plantea el riesgo de que las respuestas regulatorias no sean adecuadas para resolver las preocupaciones y pongan en riesgo a usuarios inocentes, incluidos los periodistas y las comunidades vulnerables.

### **RIESGOS DE LA AUTORREGULACIÓN PARA LOS DERECHOS HUMANOS**

Las decisiones que toman las plataformas sobre qué tipos de contenido están permitidos y cómo se presentan a los usuarios afectan la capacidad de estos para expresar sus ideas y acceder a información en línea. La moderación y la curación de contenido tienen un efecto tanto a nivel individual como colectivo, ya que las decisiones individuales tomadas por plataformas —en especial aquellas que son plataformas dominantes a escala global— tienen un impacto acumulativo, moldeando el espacio de discusión y potencialmente silenciando las voces de comunidades enteras. Esto puede representar un riesgo significativo para los usuarios en general, y para comunidades vulnerables y marginalizadas en particular. A continuación, exponemos algunos de los riesgos más prominentes que se desprenden de las prácticas de moderación y curación de contenido.



## **RIESGOS DE LA MODERACIÓN DE CONTENIDO**

### **FALTA DE CLARIDAD Y TRANSPARENCIA EN LAS NORMAS**

No siempre queda claro cómo las plataformas en línea toman las decisiones de moderación de contenido. Investigaciones recientes muestran que las bajas de contenido son, en mayor parte, arbitrarias y están sujetas al criterio de las plataformas.<sup>28</sup> Esto es preocupante especialmente en un contexto en el que los gobiernos ejercen presión pública sobre las plataformas para que moderen contenido cada vez más rápido y con más precisión, sobrepasando ampliamente la capacidad técnica real para hacerlo. Se espera que existan incoherencias, ya que quienes toman las decisiones de moderación de contenido trabajan en plazos muy ajustados y con poca capacitación o apoyo,<sup>29</sup> y, tradicionalmente, carecen de diversidad geográfica, lingüística y cultural. Cuando la moderación está asistida por computadora, las decisiones automatizadas pueden fallar terriblemente en la comprensión de las sutilezas contextuales del lenguaje.<sup>30</sup>

La falta de transparencia significativa pone en peligro el goce de los derechos de los usuarios a apelar y acceder a reparaciones adecuadas. Además, la falta de transparencia y mecanismos adecuados de rendición de cuentas afectan negativamente los derechos de los usuarios a recibir e impartir información. Este es un problema apremiante, ya que Internet se ha convertido en una herramienta esencial para la participación en actividades y la discusión de temas de interés público.<sup>31</sup> Por lo tanto, las plataformas en línea que funcionan como "plazas públicas modernas"<sup>32</sup> participan directamente en la formación del discurso público.

### **FALTA DE DERECHO A LA DEFENSA Y RECURSOS EN LAS DECISIONES**

Frecuentemente, los usuarios tienen pocas oportunidades —o ninguna— de responder a la baja de contenido, confirmar su legitimidad u obtener medidas correctivas ante una remoción indebida. Existen casos conocidos en los que contenido con valor artístico o histórico ha sido dado de baja en virtud de interpretaciones demasiado estrictas de los términos de servicio de una empresa.<sup>33</sup> Por estas razones, el relator especial de la ONU sobre la promoción y protección del derecho a la libertad de opinión y expresión exigió, al igual que muchas organizaciones de la sociedad civil, mayor transparencia y rendición de cuentas en las decisiones de moderación de contenido.<sup>34</sup> Quizás como resultado de ello, algunas plataformas —incluyendo a actores dominantes— han comenzado a compartir más información sobre sus procesos internos a fin de brindar más claridad sobre sus mecanismos de toma de decisiones de moderación de contenido.<sup>35</sup>

### **MODERACIÓN DE CONTENIDO POR PARTE DE PROVEEDORES DE INFRAESTRUCTURA**

Las decisiones de moderación de contenido tomadas por intermediarios que operan a nivel de infraestructura (como los servicios de seguridad en la red y en la nube) suscitan preocupaciones adicionales. Sus decisiones, especialmente si la infraestructura de Internet sigue consolidándose, pueden terminar imposibilitando el acceso a sitios web o servicios completos. Esa es una medida extrema que debe ser considerada y evaluada detenidamente, teniendo en cuenta normas y principios claros de necesidad y proporcionalidad al igual que las consideraciones que tienen los estados al ordenar bloqueos de sitios o servicios completos.<sup>36</sup>



## **RIESGOS DE LA CURACIÓN DE CONTENIDO**

### **LOS RIESGOS DE LA "ECONOMÍA DE LA ATENCIÓN"**

El modelo de negocio de algunas plataformas en línea depende en gran parte de la atención y participación del usuario, consideradas y tratadas como un recurso económico. El tiempo que pasan los usuarios en las plataformas en línea es uno de los factores clave que determinan las ganancias económicas de la plataforma mediante la muestra de publicidades.

La mayoría de las plataformas en línea curan sus páginas de inicio y los resultados de búsqueda para aumentar la relevancia y la interacción, y así obtener ganancias de publicidades segmentadas basadas en gran medida en formas de seguimiento en línea que invaden la privacidad. Esto puede incentivar la amplificación de información sensacionalista, homogénea y de baja calidad.

La "economía de la atención" y la competencia en los "mercados de la atención" han tomado desprevenidos a los reguladores.<sup>37</sup> Los actuales modelos regulatorios gubernamentales no han sido capaces de abordar de manera efectiva los problemas impuestos por formas de recomendación de información que son invasivas, no transparentes y no consensuadas —como las "cronologías" y las "páginas de inicio" de servicios de redes sociales— que pueden afectar el derecho a formar y expresar opiniones en línea.

### **RIESGOS PARA LA LIBERTAD DE OPINIÓN**

La libertad de sostener opiniones sin interferencias es un derecho absoluto. Esto significa que no admite excepciones ni restricciones y supone la libertad de mantener creencias, ideologías, reacciones y posiciones sin coerción alguna.<sup>38</sup>

La manera en que las plataformas en línea diseñan sus interfaces y curan los flujos de información puede acarrear consecuencias para el derecho a la libertad de opinión. Las plataformas deciden qué información se presenta de manera más prominente en las interfaces de los usuarios, combinando contenido orgánico y patrocinado de maneras que no siempre son transparentes o que no habilitan la elección del usuario. También pueden influenciar la formación de opiniones las maneras que usan las plataformas para permitir y, a veces, incentivar las reacciones mediante "empujoncitos" o "nudges". La mayoría de estos mecanismos carece del escrutinio de los usuarios y del público en general, incluidos los investigadores y reguladores.

### **FALTA DE INFORMACIÓN ACERCA DE LOS MECANISMOS DE CURACIÓN**

Usualmente, los usuarios no son conscientes ni cuentan con la información suficiente acerca de cómo se jerarquiza el contenido mediante algoritmos de recomendación e información. Algunos investigadores descubrieron que los algoritmos de recomendación pueden llevar a que los usuarios lleguen a contenido cada vez más radicalizado y perturbador.<sup>39</sup> En otro caso muy controvertido, se determinó que Facebook manipulaba las cronologías de los usuarios para llevar a cabo experimentos sobre el efecto de la curación de contenido en las emociones humanas.<sup>40</sup>

La falta de transparencia impide que los usuarios, los reguladores y el público en general descubran y aborden problemas como los mencionados anteriormente. La transparencia es esencial para habilitar auditorías independientes y evitar resultados no deseados en la curación automatizada y la publicidad segmentada, lo que, en el contexto de las plataformas,

puede provocar efectos discriminatorios.<sup>41</sup> Además, la transparencia en los mecanismos de curación es una precondition para la verdadera libertad de acción del usuario sobre las herramientas que forman su paisaje informativo.<sup>42</sup>

## **PREOCUPACIONES SOBRE LA PROTECCIÓN DE DATOS**

La curación de contenido realizada para maximizar los ingresos por publicidad está basada en la recopilación de datos generalizada, llevada a cabo comúnmente sin la debida transparencia, el consentimiento del usuario, ni otras salvaguardas básicas para la protección de datos. La curación debería tener lugar en un contexto de estricto cumplimiento de las leyes de protección de datos, como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea. Los países que aún no cuentan con marcos de protección de datos y de privacidad adecuados deben priorizar la adopción de estos para poder instaurar los mecanismos de protección más altos para el derecho a la privacidad.<sup>43</sup>

## **RIESGOS DE LA CORREGULACIÓN PARA LOS DERECHOS HUMANOS**

Los modelos corregulatorios son aquellos que combinan elementos de la regulación estatal y de la autorregulación, y pueden estar basados en regulaciones o establecerse mediante acuerdos voluntarios. Tradicionalmente, la corregulación supone la cooperación formal entre actores estatales y privados en instituciones conjuntas. Estos modelos pueden representar una ventajosa forma de gobernanza, siempre y cuando sean transparentes, realmente participativos, se formulen en base a investigación y evidencia, estén sujetos a una supervisión democrática efectiva y los acompañen mecanismos de reparación adecuados.

Algunos mecanismos de corregulación adoptados recientemente, como el Código de Conducta para la Lucha contra la Incitación Ilegal al Odio en Internet de la UE, no cumplen con estas importantes condiciones.<sup>44</sup>

## **FALTA DE TRANSPARENCIA SIGNIFICATIVA Y ROBUSTA Y DE PARTICIPACIÓN PÚBLICA OPORTUNA**

Algunas iniciativas de corregulación se incorporan mediante negociaciones directas entre los gobiernos y las plataformas, de maneras que no son suficientemente transparentes y no permiten la participación oportuna de las partes interesadas o del público en general. Asimismo, las entidades de gobierno a cargo de diseñar los compromisos en algunos de esos acuerdos generalmente no son miembros de los parlamentos, y no siempre someten los acuerdos a la posterior aprobación de legisladores.

Esto puede causar problemas de rendición de cuentas, ya que no es posible que el público en general y los sujetos de los gobiernos involucrados revisen las negociaciones y examinen los intereses, las propuestas, y las concesiones de cada parte. Este fue el caso de la adopción inicial del "Llamamiento de ChristChurch"<sup>45</sup> que mencionamos anteriormente y del primer informe de transparencia del Foro Global de Internet contra el Terrorismo (GIFCT).<sup>46</sup> A la fecha de publicación de este informe, el GIFCT y el Llamamiento de ChristChurch están atravesando un proceso de reforma que apunta a resolver algunos de estos problemas, lo que representa un paso en la dirección correcta.

## **INICIATIVAS APRESURADAS Y FALTA DE CERTEZA LEGAL**

Al igual que con la regulación estatal, los esfuerzos de corregulación también pueden hacerse apresuradamente en el intento de brindar soluciones fáciles y a corto plazo a fenómenos altamente complejos. A menudo, las iniciativas de corregulación están diseñadas informalmente, sin una adecuada base en los marcos legales vinculantes, y crean una situación de incertidumbre legal que es altamente problemática para todas las partes, en especial para los usuarios.

Las recientes iniciativas de corregulación de la UE son ejemplos de políticas apresuradas que resultan en el traslado de más presiones y responsabilidades a las plataformas en línea. Estas iniciativas incluyen el Código de Conducta para la Lucha contra la Incitación Ilegal al Odio en Internet de la UE, la Recomendación sobre las medidas destinadas a luchar, de manera eficaz, contra el discurso de odio ilegal en línea,<sup>47</sup> las Directrices relativas a la Libertad de Expresión en Internet y fuera de Internet,<sup>48</sup> y el Foro de Internet de la UE.<sup>49</sup>

Estas medidas corregulatorias no vinculantes han hecho que, gradualmente, las plataformas tengan cada vez más responsabilidades sobre el contenido ilegal generado por los usuarios. Si bien son de carácter voluntario, actúan como mecanismos para presionar a los actores privados a hacer "algo" respecto de la propagación de contenido ilegal bajo la amenaza de regulación vinculante si no cumplen con las expectativas de los encargados de formular políticas públicas. La falta de certeza legal puede provocar un cumplimiento excesivo, que en la práctica se traduce a bajas ilegítimas.<sup>50</sup>



# Guía: 26 recomendaciones para una gobernanza de contenido respetuosa de los derechos humanos

## **LAS RESPONSABILIDADES DE DERECHOS HUMANOS DE LAS PLATAFORMAS EN LÍNEA A GRAN ESCALA**

Las decisiones que toman las plataformas y los estados para regular el contenido afectan el derecho a la libertad de expresión en distintas medidas según las características del servicio en cuestión y el contexto más amplio. Por ejemplo, en un foro de debate pequeño o de acceso limitado que aloja un par de cientos de personas en debates sobre temas específicos, estas decisiones pueden interferir con la capacidad de esas personas de expresar una opinión y acceder a la información en dicho espacio.

A gran escala, la capacidad de interferencia se multiplica: una plataforma de redes sociales dominante puede dejar fuera de un debate compartido a poblaciones enteras, en un contexto en el que el acceso a servicios alternativos o de la competencia puede ser limitado. Esto supone obstáculos adicionales para el ejercicio de los derechos, no solo para la libertad de expresión y acceso a la información, sino también para derechos como la libertad de asociación, afectando negativamente la participación en la vida política y otros derechos culturales y sociales.<sup>51</sup>

Decision-makers must consider the characteristics, functioning, and context of platforms to identify the challenges to user rights derived from their use. Those in the public and private sector should derive conclusions and make decisions based on careful, timely, and participatory study.

## **IMPACTO DERIVADO DEL DOMINIO DEL MERCADO/LA ESCALA**

Algunas plataformas brindan un espacio para que dos lados de un mercado se pongan en contacto o para que los usuarios suban o intercambien contenido creativo. En muchos casos, cuando una plataforma en línea no hace nada para influenciar el intercambio de contenido, se las trata y considera como simples intermediarios. Estas plataformas se han beneficiado de las protecciones legales que limitan su responsabilidad legal del contenido que crean y suben sus usuarios.

Sin embargo, las plataformas en línea más populares hoy en día no actúan como meros medios de transmisión de la comunicación en Internet. No funcionan exclusivamente como intermediarios pasivos.<sup>52</sup> Algunas tienen características que representan más que el mero alojamiento de contenido. Estas plataformas organizan y curan el contenido creado por los usuarios en base a los datos de comportamiento de los usuarios. Cuando toman acciones para incrementar la interacción del usuario, como priorizar o cuantificar la popularidad de ciertos tipos de contenido sensacionalista — incluido el contenido para campañas de desinformación y acoso— y usan algoritmos para personalizar el contenido, la práctica de la moderación de contenido se puede convertir en un bien del cual las plataformas se benefician.<sup>53</sup> El aumento en la interacción implica que los usuarios pasan más tiempo en la plataforma, comparten más datos de sí mismos, y están expuestos a más publicidades.

La manera en la que se organiza y se presenta el contenido generado por los usuarios es relevante en relación con los derechos humanos. El uso de herramientas y estrategias de curación, como los algoritmos de recomendación y la organización de contenido de manera jerárquica en las páginas de inicio, puede tener consecuencias para los derechos de los usuarios. Por lo tanto, para proteger estos derechos, es importante conocer de qué manera una empresa estructura o modifica la jerarquía de la información.

Las empresas pueden habilitar a los usuarios a tomar un rol activo e informado en la configuración de los algoritmos de recomendación, o pueden relegarlos a un rol más pasivo. Si el libre accionar del usuario está limitado y no existe una transparencia significativa, un servicio debería tener más responsabilidad en la provisión de salvaguardas en consonancia con los principios de derechos humanos.

### **IMPACTO DERIVADO DEL DOMINIO DEL MERCADO/LA ESCALA**

Sea como fuere que elijamos categorizar las plataformas en línea, aquellos que toman decisiones de gobernanza de contenido, particularmente los gobiernos al adoptar regulaciones, deben considerar el contexto actual, en el que muchas plataformas importantes que operan a nivel mundial controlan una enorme parte del mercado y toman decisiones altamente sofisticadas en cuanto al contenido. A menudo, a estas plataformas se las conoce como los guardianes de la sociedad de la información<sup>54</sup> porque ejercen un poder sustancial sobre el discurso público. También se encuentran en una posición extraordinariamente poderosa para afectar, de manera directa o indirecta, el comportamiento de los usuarios de sus servicios. En consecuencia, estas plataformas deberían asumir un mayor grado de responsabilidad hacia sus usuarios y el público en general.

56

21

### **IMPACTO DERIVADO DE LA UTILIDAD TÉCNICA DE LAS PLATAFORMAS Y LOS SERVICIOS**

Además de las plataformas de redes sociales, existen otros intermediarios de comunicaciones cuyas decisiones tienen un enorme poder sobre el discurso público. Algunos ejemplos incluyen a los proveedores de servicio de Internet, los servicios de seguridad de la red y los proveedores de almacenamiento, entre otros. Estos controlan partes importantes de la infraestructura de las comunicaciones, y a menudo forman parte de los servicios que pueden elegir los usuarios y propietarios de sitios. Por ejemplo, la decisión del proveedor de seguridad en la red Cloudflare de dejar de prestar servicio a ciertos sitios suscitó preocupaciones sobre la falta de normas claras para el caso y las dificultades de transparencia y rendición de cuentas de tales decisiones.<sup>55</sup>

El impacto que puede tener una plataforma o un servicio en el discurso público y los derechos de los usuarios varía según cómo estos funcionen (por ejemplo, si curan y organizan la visualización de la información activamente); sus capacidades técnicas (por ejemplo, si llevan a cabo funciones técnicas importantes); y su posición en el mercado (por ejemplo, si son actores dominantes). Los tomadores de decisiones relativas a la gobernanza de contenido —las plataformas en la autorregulación y los estados al adoptar regulaciones— deben considerar estas diferencias de impacto para brindar soluciones que sean específicas y apropiadas para abordar las preocupaciones de derechos humanos.

## A qué se refiere el dominio del mercado cuando se habla de regulación de plataformas

Identificar qué plataformas en línea se han vuelto tan importantes que exigen enfoques regulatorios especiales es clave en las políticas públicas de nuestros tiempos. Los reguladores de todo el mundo se enfrentan a este desafío para decidir sobre la necesidad y el momento oportuno para reformar sus marcos de políticas de competencia o al considerar nuevas regulaciones para las plataformas.

Diferentes autores exponen distintos elementos a tener en cuenta al momento de crear normas que definan e interpreten el dominio en línea. La siguiente es una lista no exhaustiva de elementos que los reguladores podrían incorporar en el análisis:

- ▶ cantidad de usuarios afectados / potencialmente afectados por la decisión
- ▶ grado de participación en la edición y curación de contenido (en actividades de curación de contenido)
- ▶ dominio del mercado
- ▶ estado legal / económico del servicio (p. ej.: clasificación como proveedores comunes, operación bajo marcos de licencia limitada, etc.)
- ▶ posición de control técnico sobre las comunicaciones (operadores de infraestructura y conectividad)
- ▶ costo de la exclusión<sup>56</sup>

## A. Recomendaciones para una regulación estatal respetuosa de los derechos humanos

### I. ATENERSE A ESTRICITOS PRINCIPIOS DEMOCRÁTICOS

Al establecer normas sobre las obligaciones de los estados y las responsabilidades de los intermediarios de proteger los derechos humanos de los usuarios, las autoridades nacionales deberían adoptar un marco legal formal para garantizar la certeza legal, la legitimidad y la armonización de los regímenes. Para garantizar la protección del derecho a la libertad de expresión, un instrumento jurídico debe contener salvaguardas de protección establecidas mediante un proceso democrático que respete los principios del modelo de múltiples partes interesadas y transparencia, y que se someta al debate público. Sea cual fuere la forma que tome, la ley debe ser previsible y accesible.<sup>57</sup>

La adopción de tal instrumento jurídico no excluye necesariamente el uso de modelos corregulatorios. Sin embargo, todo modelo corregulatorio debe estar fundado en las bases de un marco legal vinculante adoptado por los actores estatales, a fin de que todos los mecanismos de rendición de cuentas necesarios estén incorporados para evitar que los actores privados tomen decisiones posiblemente no transparentes y arbitrarias.

El marco legal formal debe tener un alcance claro y definido, contener la definición de procedimientos asociados —como notificación y acción— y establecer altos estándares de transparencia tanto para los estados como

para las plataformas en línea. Sobre todo, el marco legal debe enfatizar la clara distinción entre las obligaciones de los estados y las responsabilidades de los actores privados en cuanto a la protección de los derechos humanos de los usuarios.

Incluso los códigos de conducta voluntarios de naturaleza corregulatoria adoptados por organizaciones internacionales y negociados directamente con actores privados, como el Código de Conducta para la Lucha contra la Incitación Ilegal al Odio en Internet de la UE, deben cumplir con un marco legal establecido.

Normalmente, bajo la regulación estatal, las autoridades judiciales independientes e imparciales y otras autoridades competentes tienen la responsabilidad de implementar y supervisar

los regímenes legales o regulatorios adoptados. Todo marco legal que apunte a la regulación de interferencias con la diseminación de contenido debe cumplir con el principio de certeza legal. La legislación aplicable a los intermediarios de Internet y los usuarios en línea debe ser accesible por parte de todos los actores que estén alcanzados por ella, incluidas todas las partes afectadas. La ley debe ser previsible para que todas las personas tengan plena consciencia de sus derechos y obligaciones.

### Qué proponemos

Toda forma de regulación estatal debería estar basada en un marco legal formal adoptado mediante un proceso transparente, inclusivo y democrático. La regulación debe tener un alcance definido con precisión que sea previsible y accesible para todos los actores, incluidos los usuarios. Debe establecer definiciones cuidadosamente equilibradas de cada procedimiento y altos estándares de transparencia. Debe contener una clara distinción entre las obligaciones de los estados y las responsabilidades de los actores privados para proteger los derechos humanos de los usuarios.

## 2. ADOPTAR MECANISMOS DE PUERTO SEGURO Y EXENCIONES DE RESPONSABILIDAD LEGAL

Los regímenes de responsabilidad legal de los intermediarios son una parte fundamental de toda regulación estatal relativa a la gobernanza de contenido. Su objetivo ha sido establecer un equilibrio entre posibilitar y apoyar la innovación en el mercado de los servicios digitales y fortalecer la protección del derecho a la libertad de expresión y acceso a la información de los usuarios de Internet.

Si bien los regímenes de responsabilidad legal de los intermediarios existen en jurisdicciones nacionales de todo el mundo, los modelos de EE. UU. y de la UE han sido cruciales para la formulación de otros modelos legales nacionales. La Sección 230 de la Ley de Decencia en las Comunicaciones (CDA) les brinda inmunidad de responsabilidad legal a las plataformas de Internet respecto de la mayor parte del contenido generado por los usuarios en EE. UU., incluida la remoción de discurso lícito de las plataformas en línea y llevada a cabo por estas. Por lo tanto, es la disposición de puerto seguro más fuerte en el mundo. La segunda ley estadounidense que sirvió de inspiración para los legisladores de la UE al momento de crear su propio



marco de responsabilidad legal de intermediarios es la Ley de Derechos de Autor para Medios Digitales en el Nuevo Milenio (DMCA). Entre otras características, la DMCA incorpora limitaciones sobre la responsabilidad legal de derechos de autor para los proveedores de servicio de Internet cuando actúan como meros medios de transmisión del material infractor. Por último, la Directiva europea sobre el Comercio Electrónico prevé la exención de responsabilidad legal para los intermediarios si solo actúan como medios de transmisión de la información producida por los usuarios. La justificación detrás de estos regímenes es doble: fortalecen la protección del derecho a la libertad de expresión en línea y, al mismo tiempo, destraban el poder innovador del ecosistema de Internet. Se puede afirmar que estas leyes, y otras leyes similares alrededor del mundo, abrieron camino a la Internet de hoy en día.

La protección de las plataformas en línea otorgada por la Sección 230 de la CDA es única y estuvo influenciada en gran parte por el amplio alcance protector del libre discurso proporcionado por la Primera Enmienda de la Constitución de EE. UU. Según la interpretación de la Sección 230 de los tribunales estadounidenses, la cláusula de puerto seguro excluye no solo la responsabilidad legal estricta de las plataformas, sino también la responsabilidad legal de intermediarios de los distribuidores, como los operadores de sitios web.<sup>58</sup> El principal objetivo de esta cláusula es otorgar inmunidad reglamentaria a los intermediarios en línea, sin perjuicio del tipo de servicio que presten. En los últimos años, ambos modelos regulatorios fueron sometidos a un mayor escrutinio. Algunos afirman que debilitan las protecciones de los usuarios en línea. En EE. UU., recientes interpretaciones de los tribunales resultaron en una total inmunidad para las plataformas, dejando sin reparaciones a las víctimas de irregularidades.<sup>59</sup>

La Directiva europea sobre el Comercio Electrónico, por su parte, apunta a armonizar los estándares mínimos para la responsabilidad legal de los intermediarios en todos los estados miembros de la UE. El régimen de exención de responsabilidad legal en virtud de los artículos del 12 al 15 difiere de la cláusula de puerto seguro de EE. UU. Según el régimen de la UE, cuando un intermediario toma conocimiento efectivo de que existe contenido ilegal en su plataforma, debe actuar diligentemente o asumir la responsabilidad legal. Para habilitar esta atribución de responsabilidad legal, el modelo condicional europeo de responsabilidad legal está directamente vinculado a un procedimiento de notificación y acción que no está previsto por ley. Su implementación queda completamente a criterio de los estados miembros de la UE.

La principal diferencia entre la disposición de puerto seguro en la Sección 230 y el enfoque europeo es que la primera garantiza que los intermediarios no serán considerados responsables ante la ley si editan, filtran o eliminan contenido, ni siquiera si el contenido goza de la protección de la Primera Enmienda. El marco europeo actual carece de una protección de puerto seguro para las plataformas en línea que buscan abordar el contenido ilegal y dañino de manera más proactiva, incluso si se ven presionados a hacerlo por actores estatales y formuladores de políticas públicas a través de medidas de corregulación no vinculantes. Dicha carencia de certeza legal incentiva la remoción y el cumplimiento excesivos para satisfacer los deseos de los formuladores de políticas públicas a fin de escapar a la amenaza de la responsabilidad legal.



## Qué proponemos

Apoyamos las cláusulas de puerto seguro que permiten la moderación de contenido respetuosa de los derechos humanos. Estas cláusulas deben establecerse mediante normas claras formuladas por un marco legal formal. No son aceptables las medidas proactivas "voluntarias" impuestas sobre los intermediarios mediante presiones estatales. La delegación de las funciones gubernamentales en la autorregulación no debería estar permitida en una sociedad democrática bajo ninguna circunstancia.

- ▶ Los intermediarios deberían estar protegidos de la responsabilidad legal del contenido de terceros mediante un régimen de puerto seguro. Sin embargo, nos oponemos a la inmunidad completa para los intermediarios, ya que les impide asumir cualquier tipo de responsabilidad, lo que deja a las víctimas de irregularidades sin respaldo, sin acceso a la justicia, ni mecanismos de apelación.
- ▶ Para fortalecer el principio de certeza legal y predictibilidad para los intermediarios y los usuarios en línea, las normas que protejan a los intermediarios deben ser claras y precisas y, al mismo tiempo, dar lugar a maneras de lidiar con el contenido ilegal cuando sea de una ilegalidad manifiesta o cuando el intermediario reciba una notificación sobre su ilegalidad.
- ▶ Los regímenes de responsabilidad legal estrictos nunca son apropiados para lidiar con el contenido ilegal en línea, debido a que pueden incentivar a las plataformas a monitorear el contenido excesivamente.

## 3. NO IMPONER UNA OBLIGACIÓN DE MONITOREO GENERAL

Una obligación de monitoreo general, impuesta por actores estatales a los intermediarios, es la orden de llevar a cabo un monitoreo activo del contenido y la información que comparten los usuarios, usualmente a través de medidas automatizadas para el reconocimiento de contenido, aplicadas de manera indiscriminada y por un período ilimitado.<sup>60</sup> Este tipo de monitoreo vulnera el derecho a la libertad de expresión y, por lo tanto, nunca debería imponerse a las plataformas en línea.

Según los Principios de Manila,<sup>61</sup> no se debería exigir nunca que los intermediarios monitoreen el contenido de manera proactiva como parte de un régimen de responsabilidad legal de intermediarios. La Recomendación del Consejo de Europa sobre las funciones y responsabilidades de los intermediarios de Internet<sup>62</sup> establece que las autoridades estatales "no deberían imponer directa o indirectamente una obligación general a los intermediarios de controlar el contenido al que dan acceso o que transmiten o almacenan, sea por medios automatizados o no". En el informe de 2018 sobre la promoción y protección del derecho a la libertad de opinión y expresión,<sup>63</sup> el Relator Especial de la ONU, David Kaye, aclara que los estados deben abstenerse de establecer leyes o acuerdos que requieran el monitoreo o filtrado "proactivo" de contenido, ya que hacerlo entraría en conflicto con el derecho a la privacidad y, probablemente, resultaría en una censura previa a la publicación. Sin embargo, el monitoreo general voluntario ejercido por actores privados bajo su propio criterio no está actualmente prohibido por ninguna legislación.

El régimen legal actual que regula la responsabilidad legal de los intermediarios respecto del contenido generado por los usuarios en Europa prohíbe a los estados imponer una obligación de monitoreo general del contenido generado por los usuarios a las plataformas en línea. Esta disposición, junto con el modelo condicional de responsabilidad legal en el Artículo 14, busca fortalecer la protección del derecho a la libertad de expresión y a la información de cada usuario en línea.

### **LA DIFERENCIA ENTRE OBLIGACIONES DE MONITOREO GENERAL Y MONITOREO ESPECÍFICO**

En la práctica, resulta casi imposible brindar una clara distinción entre el monitoreo general y el monitoreo específico del contenido en línea. Sobre la base de la sentencia del Tribunal de Justicia de la Unión Europea (TJUE) del caso *L’Oreal y otros*,<sup>64</sup> el monitoreo específico es la conducta ejercida por una plataforma que involucra vulneraciones de la misma naturaleza cometidas por el mismo actor y del mismo derecho, y solo por un período estrictamente determinado y limitado.<sup>65</sup> Además, el monitoreo específico debe estar dirigido a un caso específico; es decir, debe estar limitado en términos de sujeto y duración del monitoreo.<sup>66</sup> Si bien la jurisprudencia mencionada anteriormente involucra la vulneración del derecho de autor, la técnica de filtrado aplicada sigue siendo la misma, independientemente de los diferentes tipos de contenido abordado.

En la práctica, el monitoreo específico del contenido en línea se llevaría a cabo mediante el uso de herramientas automatizadas, como las tecnologías de reconocimiento de contenido. Para ser eficaces, estas tecnologías tendrían que aplicarse a todo el contenido generado por los usuarios que esté alojado en las plataformas en línea, sin importar los diferentes contextos del contenido en cuestión. Por lo tanto, el monitoreo específico puede terminar imponiendo la obligación de que las plataformas en línea prevengan la carga de contenido ilegal y monitoreen todo el contenido de manera activa en sus plataformas para lograr ese fin.<sup>67</sup> Esto daría paso a la evasión de la prohibición de monitoreo general que existe actualmente en virtud de la legislación europea.

Por lo tanto, considerando que el monitoreo específico puede convertirse fácilmente en monitoreo general, el primero no debería estar formulado como una obligación legal impuesta a los intermediarios por un marco legal formal.

#### **Qué proponemos**

No Los estados no deberían imponer obligaciones de monitoreo general ni específico a los actores privados.

Todo uso de tecnologías de reconocimiento de contenido que resulte en el ejercicio indirecto de monitoreo general no debería estar ordenado ni permitido por ley.

Ninguna disposición jurídica debería ordenar, incentivar o darles a las plataformas ninguna indicación de que deberían filtrar contenido de manera proactiva antes de que este sea publicado.

## **4. DEFINIR MECANISMOS DE RESPUESTA ADECUADOS**

Muy pocos países especifican los procedimientos de notificación y acción en sus leyes nacionales. La falta de armonización en los procedimientos ha

causado disparidades en la implementación de regímenes de responsabilidad legal de intermediarios en Europa y otras regiones. Para garantizar respuestas adecuadas que estén adaptadas a la categoría específica de contenido generado por los usuarios, un marco legal formal debe establecer procedimientos específicos para los mecanismos de notificación.

Dichos mecanismos deben ser visibles, fácilmente accesibles y amigables con los usuarios, y también estar adaptados al contexto. Un ejemplo de malas prácticas en este sentido es el cumplimiento de Facebook de la Ley de Aplicación de Redes de Alemania (NetzDG). El formulario especial de notificación exigido por la ley alemana, como lo emplea Facebook, es apenas visible para sus usuarios y el diseño de la interface está mal logrado, especialmente en comparación con el mecanismo diseñado por Facebook para alertar sobre contenido que, supuestamente, viola las Normas Comunitarias de Facebook, que es más claro y más fácil de usar.

### LOS DIFERENTES TIPOS DE CONTENIDO DEBERÍAN EXIGIR DISTINTAS FORMAS DE NOTIFICACIÓN

Si el contenido es de acceso público (es decir, es visible incluso para aquellos que no están suscriptos), entonces debería ser posible que las personas que no están registradas presenten una notificación. Es importante que los mecanismos de notificación y acción especialmente diseñados para categorías determinadas de contenido o actividades ilegales en línea estén acompañados de salvaguardas de derechos humanos que limiten posibles injerencias. Los procedimientos de notificación y acción diseñados especialmente permiten una mejor mitigación de los conflictos de derechos humanos en los casos que involucran la responsabilidad legal de los intermediarios.

Para que los mecanismos de notificación sean más efectivos, deberían ser fáciles de usar y maximizar la información que se les da a los usuarios. Por lo tanto, las plataformas deberían brindar una lista de razones por las cuales se presenta la notificación.

## 5. ESTABLECER NORMAS CLARAS PARA CUANDO NO APLIQUEN LAS EXENCIONES DE RESPONSABILIDAD LEGAL

Como hemos explicado, en ciertos marcos legales, cuando la plataforma toma "conocimiento efectivo" de un contenido infractor, se la despoja de sus protecciones de puerto seguro. La ley debe establecer estándares claros para determinar cuándo y cómo los intermediarios toman "conocimiento efectivo" del contenido ilegal en sus plataformas.

### Qué proponemos

En la mayoría de los casos, solo las órdenes emitidas por un tribunal o un organismo administrativo imparcial e independiente pueden constituir conocimiento efectivo de la ilegalidad del contenido de un tercero. Esto requiere una reacción expedita por parte de los intermediarios de comunicaciones para que puedan conservar su inmunidad legal.

Como estipula la Recomendación del Consejo de Europa sobre las funciones y responsabilidades de los intermediarios de Internet, una notificación privada puede llegar a considerarse como toma de conocimiento efectivo solo en el caso del contenido de ilegalidad manifiesta y siempre y cuando:

- ▶ el contenido sea de una ilegalidad manifiesta
- ▶ la mayoría de los intermediarios pueda reconocer e identificar el contenido de ilegalidad manifiesta
- ▶ el tipo de contenido esté definido claramente por ley y la notificación de su presencia en la plataforma no conduzca a ningún tipo de obligación de monitoreo proactivo.

Sin embargo, los intermediarios no deberían estar obligados ni motivados a buscar contenido ilegal de manera activa bajo ninguna circunstancia.

## **6. EVALUAR DETENIDAMENTE Y DE FORMA LIMITADA EL CONTENIDO DE ILEGALIDAD MANIFIESTA**

Un marco legal debe especificar qué tipo de contenido es de ilegalidad manifiesta y qué grupos vulnerables deberían recibir protección especial (por ejemplo, los menores de edad). El contenido es de ilegalidad manifiesta cuando es fácilmente reconocible como tal sin necesidad de mayor análisis factual o legislativo por parte de las plataformas.<sup>68</sup> Un ejemplo típico de dicho contenido es el material de explotación infantil en Internet.

La lista del contenido de ilegalidad manifiesta debería estar determinada por ley y ser específica, refiriéndose a definiciones concretas en los códigos penales nacionales, disposiciones del derecho civil, y otras medidas jurídicas relevantes. Debido a que la mayor parte del contenido ilegal requiere un análisis del contexto, los actores y los comportamientos, los casos de contenido de ilegalidad manifiesta deberían estar definidos y ser interpretados de manera restringida. Principalmente, toda la legislación relevante, especialmente las disposiciones del derecho penal nacional, debería ser sometida a una revisión para cumplir con los estándares internacionales de derechos humanos estipulados en el Pacto Internacional de Derechos Civiles y Políticos (PIDCP)<sup>69</sup> y el Convenio Europeo de Derechos Humanos (CEDH).<sup>70</sup>

En el caso del contenido de ilegalidad manifiesta, la ley podría exigir que las plataformas restrinjan temporalmente el acceso al contenido al tomar conocimiento mediante una notificación privada, antes de recibir la orden de un tribunal o un organismo administrativo independiente de eliminar el contenido. Un funcionario independiente debería llevar a cabo, con inmediatez, una revisión de cada uno de los casos.

### **Qué proponemos**

No eliminar adecuadamente el contenido que se considera de ilegalidad manifiesta tras una notificación privada presentada por un tercero debería ser la única situación en la que las plataformas puedan ser responsabilizadas legalmente por no eliminar contenido sin que se lo haya ordenado un funcionario independiente.

Las plataformas nunca deberían ser responsabilizadas legalmente por no eliminar el contenido que no es de ilegalidad manifiesta según las disposiciones jurídicas específicas.

El abuso de las notificaciones privadas debería estar desaconsejado en la regulación, estableciendo sanciones apropiadas y proporcionadas contra los usuarios o los alertadores de confianza que estén involucrados.

## 7. CREAR PROCEDIMIENTOS DE NOTIFICACIÓN Y ACCIÓN RESPETUOSOS DE LOS DERECHOS FUNDAMENTALES

Los procedimientos de notificación y acción son mecanismos que siguen las plataformas en línea con el propósito de combatir contenido ilegal al recibir una notificación.<sup>71</sup> "Notificación y acción" es un término amplio que comprende varios mecanismos con distintos tipos de respuestas ante el contenido ilegal. Sin embargo, todos inician con una notificación.<sup>72</sup> Los marcos legales formales deben establecer los mecanismos de notificación y acción más propicios.

Los diferentes tipos de contenido y actividad ilegal en línea pueden requerir distintas respuestas, especialmente diseñadas para el tipo de contenido generado por los usuarios que se quiera abordar. No obstante, la ley debe definir de manera clara los procedimientos y brindar salvaguardas apropiadas para su aplicación por parte de los estados.

Este informe propone los siguientes procedimientos de notificación y acción para lidiar con tipos específicos de contenido generado por los usuarios. La propuesta surge de investigaciones profundas llevadas a cabo por distintos expertos en el campo de la gobernanza de contenido.<sup>73</sup>

En primer lugar, para la violación de derechos de autor, los mecanismos de notificación y notificación han demostrado ser la medida más equilibrada para abordar el material de derecho de autor compartido en las plataformas.<sup>74</sup> Mediante el mecanismo de notificación y notificación, un intermediario recibe una notificación con una queja, la cual luego se envía al proveedor de contenido.<sup>75</sup> Allí termina la participación del intermediario. Una vez que el proveedor de contenido recibe la información de la plataforma, el caso queda en manos del supuesto principal infractor; es decir, el usuario. Luego, la parte notificada debe decidir entre eliminar el contenido o responder la notificación dentro del período prescripto.<sup>76</sup>

En segundo término, para el contenido generado por los usuarios que depende del contexto en una mayor medida y específicamente para los supuestos casos de difamación, sugerimos implementar un mecanismo de notificación, espera y baja. Mediante este procedimiento, los intermediarios deben reenviar toda notificación concerniente a contenido supuestamente difamatorio e ilegal a ese proveedor de contenido y luego esperar una semana antes de imponer el bloqueo o la remoción. Esta "interpretación menos estricta de la notificación y baja"<sup>77</sup> garantiza al proveedor de contenido la posibilidad de emitir una respuesta antes de que se tome una acción. Por lo tanto, el mecanismo permite que se impidan las bajas indebidas de contenido y les da a los usuarios la oportunidad de ser escuchados.

En todas las circunstancias, la notificación y baja judicial, que se refiere a la orden de bloquear o eliminar el contenido generado por los usuarios emitida por una autoridad judicial, debe siempre estar disponible para todos los usuarios en línea, incluidos los proveedores de contenido y las víctimas de injerencias.

Idealmente, se deberían incorporar salvaguardas —como las del recuadro a continuación— directamente en los procedimientos de notificación y acción. Para minimizar los resultados no deseados de los procedimientos, debería aplicarse un tipo de notificación y acción en particular, basado en el objetivo que se persiga y el potencial riesgo a los derechos protegidos.

## Qué proponemos

A fin de garantizar la protección de los derechos fundamentales, los detalles de los procedimientos de notificación y acción deben estar claramente definidos por ley y tener propósitos específicos.

Los estados deben evaluar con regularidad los posibles efectos no deseados de toda restricción antes y después de aplicar los determinados procedimientos de notificación y acción.

Los estados tienen la obligación de buscar las medidas menos intrusivas para los derechos humanos de los usuarios.

Los siguientes procedimientos de notificación y acción se deben considerar como adecuados, según el tipo de infracción en cuestión y la categoría del contenido.

- ▶ Mecanismo de notificación y notificación
- ▶ Mecanismo de notificación, espera y baja que permita al proveedor de contenido presentar una contradenuncia
- ▶ Mecanismo de notificación y baja judicial, por el cual los tribunales revisen la legitimidad de la remoción del contenido, siempre a disposición de todos los usuarios, sin perjuicio del tipo de contenido
- ▶ Mecanismo de notificación privada y baja, únicamente utilizado en casos limitados en los que el contenido se defina como contenido de ilegalidad manifiesta

Los estados deben abstenerse de adoptar mecanismos de notificación y baja permanente que establezcan la obligación de evitar que el contenido esté disponible en el futuro, usualmente a través de medidas automatizadas que implican un monitoreo general.

## QUÉ ES UNA NOTIFICACIÓN VÁLIDA

Para que una notificación sea válida, debe contener suficiente información para que las plataformas actúen. Debe ser precisa y estar justificada adecuadamente. Las condiciones que necesita una notificación para ser válida deben estar especificadas en la ley. De todas formas, los requisitos concretos de una notificación válida deben estar determinados según el tipo de contenido en cuestión.<sup>78</sup>

## Qué proponemos

Requisitos básicos y mínimos para que una notificación sea válida:

- ▶ Razón de la queja
- ▶ Ubicación del contenido
- ▶ Evidencia que respalde la queja
- ▶ Consideración de las limitaciones, las excepciones, y la defensa a disposición del proveedor de contenido
- ▶ Declaración de buena fe

Las notificaciones presentadas por los estados deben estar basadas en su propia evaluación de la ilegalidad del contenido notificado, en consonancia con los estándares internacionales. El lenguaje para las restricciones de contenido debe contemplar la notificación de tal restricción entregada al productor/

emisor del contenido tan pronto como sea posible, a menos que eso interfiera con las actividades de aplicación de la ley en curso. La información debe estar disponible para los usuarios que buscan acceso al contenido, según las leyes de protección de datos aplicables. Los usuarios no deberían estar obligados a identificarse al presentar la notificación y deberían brindar su información de contacto solo de manera voluntaria.

### **REQUISITOS PARA LOS PROGRAMAS DE "ALERTADORES DE CONFIANZA"**

Los alertadores de confianza son entidades con experiencia específica y estructuras especializadas para la detección e identificación de contenido ilegal en línea.<sup>79</sup> Solo los organismos independientes con conocimiento experto y específico deberían estar facultados para ser alertadores de confianza. El marco legal debería estipular claramente los criterios que se deben cumplir para ser un alertador de confianza. Adicionalmente, todo marco legal que regule la responsabilidad legal de los intermediarios debería contener salvaguardas legales apropiadas para la imparcialidad e independencia de dichos alertadores, lo que garantizaría decisiones más equilibradas.

Los alertadores deberían perder su estado si alguno de los requisitos para el ejercicio de esta función se ve vulnerado. Las condiciones para la institución de alertadores de confianza no deberían, en ningún caso, estar determinadas solamente por las plataformas privadas.

#### **Qué proponemos**

Los programas de alertadores de confianza son aceptables solo si los requisitos para serlo y las salvaguardas legales para su independencia están regulados por ley. De no ser así, las obligaciones de los estados se trasladarían, tarde o temprano, a las plataformas privadas.

## **8. LIMITAR LAS MEDIDAS TEMPORALES E INCLUIR SALVAGUARDAS**

El bloqueo temporal podría aplicarse como medida temporal en circunstancias en las que el tiempo es un factor importante en la vulneración, pero solo para tipos específicos de contenido ilegal (medidas de emergencia). En los casos en los que el contenido potencialmente ilegal es objetado, las medidas temporales permitirían el bloqueo al acceso a un contenido específico a la espera de la resolución del conflicto o una respuesta por parte del proveedor de contenido o un funcionario independiente.<sup>80</sup> Las medidas de emergencia temporales deben estar limitadas de manera estricta para evitar abusos estatales de esta herramienta que hagan que otros tipos de contenido no estén disponibles sin un procedimiento apropiado para la determinación de su ilegalidad. La mayor parte del contenido ilegal, como el contenido que vulnera el derecho de autor, por ejemplo, no es de una naturaleza urgente que requiera medidas temporales. La ley debe aclarar los casos específicos en los que deben aplicarse medidas temporales.

El marco legislativo debería determinar de manera clara todos los plazos designados para los pasos de los procedimientos de notificación y acción —incluidas las medidas temporales. Los plazos para la remoción de contenido no pueden ser demasiado cortos, ya que esto incentivaría las bajas ilegítimas de contenido lícito. Varias propuestas legislativas recientes sirven de ejemplo, como la regulación propuesta por la UE para la lucha contra la



diseminación de contenido terrorista en línea, que propone un plazo de una hora para la remoción de contenido.<sup>81</sup>

Un límite de tiempo tan corto como este no debería ser considerado proporcionado a su objetivo legítimo, incluso si se lo limita a casos excepcionales. Debido al acotado límite de tiempo, no existe una posibilidad real de que el creador del contenido pueda apelar la decisión de la plataforma y, por lo tanto, de que obtenga una reparación apropiada en caso de una baja ilegítima. En consecuencia, un límite de tiempo demasiado corto plantea graves amenazas al derecho a la libertad de expresión e información, así como también al derecho de los usuarios a un juicio justo y una reparación adecuada.

Si bien en algunas jurisdicciones se debaten plazos específicos,<sup>82</sup> o hay evidencia respaldada por investigaciones que apoye un plazo límite de una hora para el cumplimiento de la remoción del contenido. Los plazos apropiados deberían depender de los contextos regulatorios y operativos de cada jurisdicción.

### Qué proponemos

Estos son algunos factores temporales que deberían estar especificados en la ley, como lo sugiere la Recomendación del Consejo de Europa sobre las funciones y responsabilidades de los intermediarios de Internet:

- ▶ el límite de tiempo para reenviar la notificación al proveedor de contenido
- ▶ el límite de tiempo para que el proveedor de contenido presente la contranotificación
- ▶ el límite de tiempo para tomar la decisión sobre la remoción del contenido o su conservación en la plataforma
- ▶ el plazo exacto para informar las decisiones fundadas a todas las partes involucradas
- ▶ el período para iniciar una revisión judicial de la decisión de un intermediario

## 9. IMPONER SANCIONES PROPORCIONADAS PARA EL INCUMPLIMIENTO

a única situación en la que los intermediarios deberían ser responsabilizados legalmente debería ocurrir cuando no cumplan con una orden de un funcionario imparcial e independiente sobre la remoción del contenido en cuestión o cuando no eliminen un contenido de ilegalidad manifiesta, según lo defina la ley, tras recibir una notificación privada. Las sanciones impuestas a los intermediarios por el incumplimiento deben ser proporcionadas. Si no lo son, es muy probable que resulten en un cumplimiento excesivo, lo que puede perjudicar la libre expresión y el acceso a la información compartida en las plataformas en línea.<sup>83</sup>

### Qué proponemos

Las sanciones impuestas a los intermediarios por el incumplimiento de solicitudes de remoción deben ser proporcionadas. Deben estar directamente correlacionadas con la falta de cumplimiento de la remoción de contenido o de la orden de restricción.



## 10. USAR MEDIDAS AUTOMATIZADAS SOLO EN CASOS LIMITADOS

Las medidas automatizadas, como los filtros de carga o el análisis de bases de datos de funciones hash de diferentes tipos, pueden ser útiles para evaluar las grandes cantidades de información que se comparten en las plataformas en línea. Sin embargo, acarrear un problema fundamental: no son capaces de interpretar el contexto antes de tomar la decisión de bloquear o dar de baja el contenido. Esto implica un grave riesgo para la libertad de expresión y el acceso a la información en la mayoría de los casos, en los que es necesario interpretar el contexto.

Para evitar falsos positivos o bloqueos excesivos y proteger el uso legítimo de partes de contenido ilegal para el interés público, como en las noticias, el uso de las medidas automatizadas debería estar aceptado solo en casos limitados de contenido de ilegalidad manifiesta que no depende del contexto, y no debería imponerse nunca como una obligación legal para las plataformas. Un ejemplo típico de este tipo de contenido es el abuso sexual de menores. En cualquier caso, la legislación debería seguir los parámetros a continuación.

### Qué proponemos

- ▶ Todo uso de herramientas automatizadas debe estar basado en políticas claras y transparentes, incluidos los mecanismos de transparencia para la valoración independiente de su creación, funcionamiento y evaluación.
- ▶ Dicho uso debe perseguir un propósito legítimo (por ejemplo, restringir el acceso a contenido ilegal específico que no depende del contexto). El propósito legítimo debería siempre estar determinado por una autoridad judicial independiente u otro organismo administrativo independiente cuyas decisiones estén sujetas a revisión judicial.
- ▶ Las plataformas pueden hacer uso de herramientas automatizadas en consonancia con sus propias políticas —incluidos los mecanismos de detección automática— siempre y cuando sean transparentes y se ajusten a los estándares internacionales de derechos humanos.
- ▶ La aplicación de medidas automatizadas proactivas respaldada por los estados que llevan a cabo las plataformas en línea para el reconocimiento de contenido no puede considerarse como conocimiento efectivo ni, consecuentemente, resultar en responsabilidad legal.
- ▶ El uso de herramientas automatizadas no debería dar paso a una obligación de monitoreo general para las plataformas u otros servicios de comunicaciones.
- ▶ Los sistemas de baja de contenido automatizados que implementen las plataformas deben permitir la revisión humana a fin de evitar falsos positivos.

## 11. LEGISLAR SALVAGUARDAS PARA EL DEBIDO PROCESO

Las siguientes salvaguardas son de extrema importancia y deben estar incorporadas en toda regulación para garantizar el debido proceso para los creadores de contenido. Además, las salvaguardas son necesarias para brindar certeza legal, predictibilidad y proporcionalidad en las medidas, a beneficio de los proveedores de contenido y también de las plataformas y usuarios.

### NOTIFICAR A LOS PROVEEDORES DE CONTENIDO

Es absolutamente esencial notificar al proveedor de contenido antes de que se tome una acción para garantizar que los usuarios conserven su derecho a un juicio justo y una reparación adecuada. La notificación debe indicar claramente la razón de la remoción o el bloqueo, brindar una explicación precisa de los derechos que posee el proveedor de contenido y sus posibilidades de apelar la decisión u optar por una revisión judicial. La única excepción a la obligación de enviar una notificación a un proveedor de contenido se aplica en situaciones en las que dicha notificación podría obstaculizar las actividades de aplicación de la ley, como la prevención, investigación, detección y persecución de delitos penales.<sup>84</sup> Sin embargo, esta excepción debe ser proporcionada respecto de su objetivo legítimo, necesaria para lograr el fin buscado, y estar basada en la ley.

#### Qué proponemos

Los usuarios que actúan como proveedores de contenido deben ser notificados antes de que se lleve a cabo una acción contra su contenido.

La notificación debería incluir, por lo menos, lo siguiente:

- ▶ las razones para la remoción o el bloqueo;
- ▶ una explicación precisa de los derechos del proveedor de contenido;
- ▶ una explicación de las posibilidades de apelación de la decisión;
- ▶ una clara explicación de la opción de acceder a recurso judicial.

Excepción: los casos en que se ponga en peligro una investigación penal de un delito grave. Las condiciones para las exenciones deben estar definidas por la ley de manera proporcionada y ajustarse a los estándares internacionales de derechos humanos.

### ESTABLECER UN MECANISMO DE CONTRANOTIFICACIÓN

Los procedimientos de notificación y acción, si se adoptan en la regulación, deben permitir la contranotificación. La contranotificación les permite a los proveedores de contenido objetar quejas individuales relacionadas con su contenido. Esta es una precondition de equidad en todo proceso de toma de decisiones. Las excepciones al mecanismo de contranotificación deben estar definidas en el marco legal formal.

#### Qué proponemos

- ▶ Las contranotificaciones son una herramienta necesaria para el derecho a la defensa en el contexto de los procedimientos de notificación y acción.
- ▶ Los usuarios deben ser capaces de presentar una contranotificación antes de que se lleve a cabo una acción contra su contenido.

- ▶ Las condiciones para el uso de las contranotificaciones especificadas en la ley no deben ser excesivamente demandantes, ya que esto podría
- ▶ La ley debe especificar qué tipo de contenido y en qué situación se puede hacer una excepción al uso de la contranotificación.

## 12. CREAR OBLIGACIONES SIGNIFICATIVAS DE TRANSPARENCIA Y RENDICIÓN DE CUENTAS

Para poder establecer una transparencia significativa que se centre en la calidad en lugar de en la cantidad de las decisiones de gobernanza de contenido, un marco legal formal debe estipular requisitos de transparencia claros y sólidos tanto para los estados como para los actores privados. Un marco legal apropiado también prevería una evaluación precisa de las políticas existentes, su funcionamiento y su efectividad.

La transparencia es una precondition para la recopilación de evidencia sobre la implementación y el impacto de las leyes existentes. Permite a los legisladores y órganos judiciales entender mejor el campo regulatorio y aprender de errores pasados. Para evitar la legislación formulada apresuradamente, que a veces pasa por alto su propio objetivo y se vuelve potencialmente intrusiva con los derechos humanos, los reguladores deben ser capaces de monitorear cómo se cumplen los objetivos iniciales. Así, un correcto monitoreo determinará si existe la necesidad de establecer más respuestas regulatorias o si el marco legal ya aplicable es satisfactorio. Este objetivo puede ser implementado en la práctica solo si se tiene acceso a los datos completos y relevantes sobre las remociones de contenido llevadas a cabo por los estados y las entidades privadas.

### CUÁLES SON LOS REQUISITOS DE LOS INFORMES DE TRANSPARENCIA EMITIDOS POR LOS ESTADOS

Los informes de transparencia deben ser presentados tanto por los estados como por los actores privados. Existen varias autoridades públicas que emiten notificaciones de remoción de contenido ilegal o monitorean las plataformas para evitar actividades ilegales. Por ejemplo, en el caso de la Ley de Aplicación de Redes de Alemania (NetzDG), que impone la responsabilidad legal condicional del contenido ilegal generado por los usuarios en las redes sociales, es la Oficina Federal de Justicia el organismo que supervisa la aplicación de la ley. En julio de 2019, la Oficina Federal de Justicia de Alemania —*Bundesamt für Justiz* o BfJ— decidió multar a Facebook por violar los requisitos de transparencia estipulados en la ley.<sup>85</sup>

Si bien es posible que la decisión tenga un impacto positivo en la transparencia, se pone de manifiesto otro problema apremiante: la BfJ no puede ser considerada como un regulador políticamente independiente, porque responde directamente al Ministerio de Justicia. Aunque hace cumplir los requisitos de transparencia a los actores privados, no está sujeta al mismo umbral de transparencia. Puede que la independencia de un regulador no sea una preocupación grave en algunos países que respetan los principios democráticos y el estado de derecho; sin embargo, para los países que sufren retrocesos democráticos o se encuentran bajo gobiernos autoritarios, la falta de un regulador que sea independiente del estado puede causar la supresión de la disidencia y otros graves abusos a los derechos humanos.

## Qué proponemos

Para proteger el discurso democrático y el escrutinio público del liderazgo político, los estados deben poner a disposición del público y de manera regular información válida sobre lo siguiente:

- ▶ la cantidad y la naturaleza de las restricciones de contenido, así como también las categorías de datos personales que se solicitaron a los intermediarios. Los informes de transparencia estatales deben incluir todas las solicitudes sobre contenido presentadas a los intermediarios.
- ▶ los fundamentos jurídicos claros sobre los que se basaron las solicitudes, incluso aquellas basadas en tratados internacionales de asistencia jurídica mutua.
- ▶ los pasos exactos que se tomaron como resultado de sus solicitudes.

## CUÁLES SON LOS REQUISITOS DE LOS INFORMES DE TRANSPARENCIA EMITIDOS POR LAS PLATAFORMAS

Las plataformas revelan muy poca información sobre cómo se formulan y aplican las normas y los mecanismos privados para la autorregulación y la corrección. Particularmente, es "increíblemente baja" la revelación de información sobre las acciones tomadas en virtud de solicitudes privadas de remoción de contenido bajo los términos de servicio.<sup>86</sup> Los estados deberían exigir que los intermediarios revelen información precisa, simple y de lectura electrónica sobre todas las injerencias de los derechos de los usuarios a la libertad de expresión e información, el derecho a la privacidad, la protección de datos personales y otros derechos fundamentales. Esta información debe ser de fácil acceso para el público en general.

Solo mediante la combinación de informes de transparencia exhaustivos por parte de los estados y los intermediarios, los reguladores y los usuarios podrán tener una idea realista de cómo funciona la moderación de contenido en línea. Cumplir con un umbral de transparencia de esta índole construiría los cimientos de la investigación para la elaboración de políticas basadas en evidencia. Para lograrlo, los actores privados deberían adoptar un enfoque de transparencia integral y significativo para sus operaciones, compartiendo información sobre cómo desarrollan e implementan las normas para la moderación de contenido.

## Qué proponemos

Los intermediarios deben incluir en los informes de transparencia, por lo menos, la siguiente información:<sup>87</sup>

- ▶ la cantidad total de notificaciones recibidas
- ▶ el tipo de entidades que las emitieron, incluidos entes privados, organismos administrativos o tribunales
- ▶ las razones para la determinación de la legalidad del contenido o cómo este infringe los términos de servicio
- ▶ plazos concretos para el envío de la notificación al proveedor de contenido antes de que se lleve a cabo una acción, para la presentación de una contranotificación, el tiempo exacto que pasará antes de que el contenido sea restringido, y el plazo para iniciar un proceso de apelación
- ▶ la cantidad de apelaciones recibidas y cómo estas se resolvieron

## **13. GARANTIZAR LOS DERECHOS DE APELACIÓN Y REPARACIÓN EFECTIVA DE LOS USUARIOS**

La toma de decisiones sobre el contenido generado por los usuarios exige un equilibrio cuidadoso y generalmente complicado de los derechos. A fin de cuentas, los errores que resultan de una evaluación incorrecta de los casos son inevitables. Por lo tanto, establecer mecanismos de apelación claros y fácilmente accesibles es la garantía principal de equidad procesal. En este informe, distinguimos dos mecanismos: la resolución de disputas directamente provista por las plataformas y los recursos judiciales garantizados por los estados.

En primer lugar, los mecanismos de apelación provistos por los intermediarios de comunicaciones deben ser accesibles, asequibles y transparentes. Aun así, no deberían remplazar nunca un recurso judicial otorgado por los tribunales. Los mecanismos de apelación deben estar disponibles para los usuarios en línea en los casos de remoción de contenido y también cuando los intermediarios se rehúsen a cumplir o ignoren las solicitudes de remoción de los usuarios. Exploraremos los mecanismos de revisión que las plataformas deberían establecer como parte de sus responsabilidades para proteger los derechos humanos a continuación.

En segundo lugar, los recursos judiciales imparciales e independientes deben estar siempre disponibles para los usuarios en línea, especialmente cuando las resoluciones de disputas a nivel de los intermediarios son consideradas insuficientes por las partes afectadas. Aunque teóricamente los recursos judiciales son siempre una opción, es muy importante que dicha opción este otorgada por ley. Asimismo, la orden judicial debe estar disponible también en casos de remoción ilegítima de contenido, para que los proveedores de contenido puedan beneficiarse equitativamente de la posibilidad de acceder a una medida cautelar.

## **B. Recomendaciones para una autorregulación respetuosa de los derechos humanos**

Las decisiones de moderación y curación de contenido pueden tener consecuencias tanto para la libre expresión como para otros derechos fundamentales, como el derecho a la libertad de asociación, y el goce de derechos económicos, sociales y culturales.

Para prevenir, abordar o mitigar las violaciones a los derechos humanos y promover, a la vez, el goce de estos, hemos desarrollado pautas básicas en base a los derechos humanos<sup>88</sup> para la moderación de contenido, las cuales se detallan a continuación.

Toda plataforma que toma decisiones sobre el discurso de terceros debería seguir estos principios, independientemente de otras obligaciones legales. A medida que las plataformas crecen en tamaño, alcance geográfico, e influencia, actuando como intermediarios del discurso público, alejarse de ellas puede representar un mayor riesgo para los derechos humanos de los usuarios. Es fundamental que las plataformas en línea dominantes, que pueden afectar significativamente el discurso,<sup>89</sup> interpreten y sigan estos principios aún más estrictamente.

Las plataformas deben considerar a los derechos humanos desde el diseño de sus productos hasta el desarrollo y la implementación de políticas y prácticas de moderación y curación de contenido. Deben apuntar a alcanzar un entorno en línea que facilite el libre intercambio de ideas, empodere a los usuarios y proteja los derechos de las comunidades vulnerables.

## **1. PREVENIR PERJUICIOS A LOS DERECHOS HUMANOS**

Platforms must consider human rights from the design of their products through the development and implementation of content moderation and curation policies and practices. This must be aimed at achieving an online environment that furthers the free exchange of ideas, empowers users, and protects the rights of vulnerable communities.

### **Qué proponemos**

Entre otras acciones, las plataformas deberían hacer lo siguiente:

- ▶ Incorporar protecciones de derechos humanos a toda nueva política y servicio, en lugar de recurrir a un modelo que las habilite a expandirse primero y abordar los abusos luego
- ▶ Consultar a otros expertos en derechos humanos y organizaciones de la sociedad civil con regularidad, en especial antes de lanzar nuevos productos, funciones o servicios

## **2. EVALUAR EL IMPACTO**

Las plataformas también deberían llevar a cabo evaluaciones públicas, periódicas y participativas para determinar la manera en que las decisiones de moderación y curación de contenido afectan los derechos fundamentales de los usuarios y tomar las medidas necesarias para mitigar cualquier daño.

### **Qué proponemos**

Las plataformas deberían:

- ▶ Compartir información de manera proactiva con los investigadores y la sociedad civil para permitirles evaluar independientemente los impactos de las decisiones de moderación y curación de contenido en los derechos humanos
- ▶ Contribuir, incluso mediante apoyo económico, al trabajo de grupos de investigación y de la sociedad civil que lleven a cabo evaluaciones independientes
- ▶ Elaborar e incorporar abiertamente en sus operaciones protocolos de evaluación de impacto sobre derechos humanos para optimizar el trabajo de los investigadores, la sociedad civil y los reguladores

## **3. SER TRANSPARENTES**

Todos los criterios, normas, sanciones y excepciones de moderación y curación de contenido deben ser claras, específicas y predecibles, e informarse a los usuarios por adelantado.

## Qué proponemos

En particular, las plataformas deberían:

- ▶ Obtener el consentimiento válido e informado de los usuarios con respecto a las normas, los criterios, las sanciones y las Garantizar que el consentimiento pueda ser revocado de manera fácil y sin obstáculos
- ▶ interpreting and applying content moderation rules, to ensure that decisions on content are as predictable and understandable as possible
- ▶ Incluir pautas que expliquen el proceso interno de la empresa para la interpretación y aplicación de las normas de moderación de contenido, con el fin de asegurar que las decisiones sobre el contenido sean tan predecibles y comprensibles como sea posible
- ▶ Describir en detalle qué constituye una violación de las normas, cuáles son las sanciones correspondientes, cómo funcionan los procesos de apelación y cómo se aplicarán las políticas
- ▶ Poner a disposición toda la información en el idioma oficial del país donde se presta el servicio, presentándola con lenguaje simple y evitando el exceso de terminología técnica y referencias a otros documentos
- ▶ Notificar a los usuarios sobre cualquier cambio en estas normas y garantizar que sean aceptados de manera explícita por los usuarios antes de que sean aplicables
- ▶ Informar a los usuarios sobre la recopilación y el uso de sus datos y poner a disposición todos los derechos otorgados por los principios y leyes de protección de datos, en una manera que sea coherente con los más altos estándares de tal protección
- ▶ Abstenerse de mantener prácticas dirigidas a "empujar" (nudge), influenciar o manipular a los usuarios sin su conocimiento o consentimiento. El uso de la tecnología de curación de contenido, como los algoritmos de recomendación y jerarquización en las páginas de inicio, debería ser lo más claro y transparente posible

## 4. APLICAR LOS PRINCIPIOS DE NECESIDAD Y PROPORCIONALIDAD

Las sanciones que las plataformas impongan a los usuarios por violar las normas de moderación de contenido deben ser proporcionales al daño que se quiere subsanar y tener en cuenta la efectividad y el impacto en los derechos del usuario (necesidad). Las penalidades severas, como prohibirle al usuario utilizar un servicio en línea, o poner fin al servicio en páginas web o aplicaciones, deberían ser una medida de último recurso y solo aplicarse en el caso de una violación grave o luego de la reincidencia de la infracción. En todos los casos, debe llevarse a cabo una evaluación exhaustiva del impacto de la medida tomada.

## 5. CONSIDERAR EL CONTEXTO

Las plataformas no deberían aplicar las normas de moderación de contenido de manera uniforme para todos los casos. Además de recurrir a los principios de derechos humanos como base universal para tomar las decisiones de moderación de contenido, las plataformas deberían considerar también las sutiles diferencias sociales, culturales y lingüísticas tanto como sea posible.



## Qué proponemos

Para lograrlo, las plataformas deberían:

- ▶ Desarrollar y revisar las normas de curación y moderación de contenido —junto con cualquier otra pauta para su interpretación— con las contribuciones constantes de la sociedad civil, los académicos y los usuarios locales
- ▶ Invertir en recursos humanos y el desarrollo del conocimiento experto social y cultural necesario para la toma de decisiones de moderación de contenido, a fin de considerar mejor los respectivos detalles y consecuencias
- ▶ Prestar especial atención a los casos que surgen en áreas de conflicto y en situaciones que pueden afectar a las poblaciones vulnerables
- ▶ Evaluar el contexto individual al juzgar el comportamiento de usuarios en particular, cuando sea posible, teniendo en cuenta comportamientos anteriores, el cumplimiento de las sanciones y otros factores

## 6. NO TOMAR POSICIONES ARBITRARIAS NI DE INJUSTA DISCRIMINACIÓN

La aplicación de decisiones de moderación de contenido en base a los contextos y otras sutiles diferencias debería ser tan coherente, sistemática y predecible como sea posible para evitar la arbitrariedad. Las plataformas deberían prestar especial atención a cómo se implementan sus normas de moderación de contenido, ya sea por parte de los moderadores de la empresa en el uso de procesos internos o por usuarios mediante mecanismos de denuncia, para garantizar que no se ataque injustamente a comunidades marginalizadas.

## 7. PROMOVER LA TOMA DE DECISIONES CON INTERVENCIÓN HUMANA

Las plataformas en línea no solo deberían recurrir a la toma de decisiones automatizada para la moderación de contenidos.

## Qué proponemos

Si las necesidades de escala o el propio volumen de la información generada por los usuarios hacen necesario el uso de tomas de decisiones automatizadas, las plataformas en línea deberían hacer lo siguiente:

- ▶ Informar claramente a los usuarios sobre el uso de la tecnología de toma de decisiones automatizada
- ▶ Limitar el uso de la tecnología de toma de decisiones automatizada al contenido que requiera una menor interpretación para que se considere en violación de los términos de uso

En cuanto a las decisiones de moderación de contenido, las plataformas deberían también hacer lo siguiente:

- ▶ Brindar a los usuarios el derecho de solicitar una revisión humana de su caso



En el caso de la curación de contenido, las plataformas en línea deberían informar a los usuarios claramente cuándo y dónde se usará la tecnología de curación en el servicio o la aplicación. Además, deberían:

- ▶ Informar de manera clara a los usuarios acerca de los criterios que se usan para la priorización o la recomendación in a clear manner
- ▶ Permitir que los usuarios modifiquen esos criterios u opten por no estar sujetos a la curación de contenido, cuando sea posible

Tanto en la curación de contenido como en la moderación automatizada, las plataformas deberían:

- ▶ Hacer que los sistemas automáticos sean tan transparentes como sea posible
- ▶ Publicar información sobre cómo se usan estos sistemas y sobre los procedimientos detrás de su aplicación
- ▶ Poner a disposición estos sistemas para llevar a cabo auditorías independientes

## 8. CREAR MECANISMOS DE NOTIFICACIÓN Y REVISIÓN

Los intermediarios de comunicaciones deberían notificar a los usuarios cuando se toma una decisión de moderación sobre su contenido o discurso.

### Qué proponemos

Esta notificación debería contener lo siguiente:

- ▶ información adecuada sobre qué impulsó la decisión
- ▶ la norma específica que ha sido quebrantada
- ▶ cómo se interpretaron las pautas de moderación de contenido
- ▶ la acción que se tomará
- ▶ instrucciones claras para la presentación de una apelación

Una notificación también debería contener la información necesaria para solicitar una revisión de la decisión. Los mecanismos de revisión deben ser directa y fácilmente accesibles y llevarse a cabo dentro de un plazo razonable, particularmente si el contenido ha quedado como no disponible provisoriamente. Podrán ser provistos por la empresa o mediante la utilización de una entidad externa, como una junta de supervisión o un consejo para todo el sector.

### Qué proponemos

En línea con las recomendaciones del Representante Especial del Consejo de Derechos Humanos sobre los negocios y derechos humanos, los mecanismos no judiciales de presentación de quejas también deberían cumplir estos requisitos:<sup>90</sup>

**Ser legítimos:** contar con una estructura de gobernanza clara, transparente y lo suficientemente independiente para asegurar que ninguna parte en un proceso de queja particular pueda interferir con la realización justa de dicho proceso

**Ser accesibles:** ser públicos para aquellos que deseen acceder a ellos y brindar asistencia adecuada para las partes perjudicadas que puedan enfrentar obstáculos al acceso, incluidos el idioma, la alfabetización, la consciencia, las finanzas, la distancia o el miedo a represalias

**Ser predecibles:** brindar un procedimiento claro y conocido con un plazo para cada etapa y claridad sobre los tipos de procesos y resultados que puedan o no ofrecer, así como también un medio de monitoreo de la implementación de cualquier resultado

**Ser equitativos:** asegurar que las partes perjudicadas tengan acceso razonable a fuentes de información, asesoramiento y conocimiento experto necesarias para participar del proceso de presentación de quejas en términos justos y equitativos

**Ser compatibles con los derechos:** asegurar que sus resultados y reparaciones se ajusten a los estándares de derechos humanos reconocidos internacionalmente

**Ser transparentes:** brindar suficiente transparencia en los procesos y resultados para responder a las preocupaciones de interés público y asumir transparencia cuando sea posible; los mecanismos no estatales en particular deberían ser transparentes respecto de la recepción de quejas y los elementos clave de sus resultados

**Estar basados en el diálogo y la participación:** centrarse en procesos de diálogo directo o mediado para buscar soluciones convenidas, y dejar la adjudicación a mecanismos independientes terceros, ya sean judiciales o no<sup>91</sup>

## 9. PROVEER REPARACIONES

Las plataformas deberían brindar medidas correctivas eficaces para los usuarios afectados por sus políticas, productos o prácticas. Esto incluye las decisiones de moderación de contenido, en los casos en que causen daños a los usuarios, como cuando las normas se aplican errónea o excesivamente.

### Qué proponemos

Entre otras ideas, las plataformas podrían brindar medidas correctivas mediante diversas vías, según el caso. Por ejemplo:

- ▶ Restaurar el contenido eliminado en casos de remoción errónea o ilegítima
- ▶ Proporcionar el derecho a réplica, con el mismo alcance del contenido que originó la queja
- ▶ Ofrecer una explicación de la medida
- ▶ Hacer que la información no esté disponible de manera temporal
- ▶ Notificar a terceros
- ▶ Providing notice to third parties
- ▶ Emitir disculpas o correcciones
- ▶ Brindar compensaciones económicas<sup>92</sup>

## 10. INVOLUCRARSE EN LA GOBERNANZA ABIERTA

Las plataformas en línea, especialmente las dominantes, deberían crear mecanismos para la participación de sus usuarios y otras partes interesadas en la gobernanza de sus aplicaciones y servicios. Tener en cuenta los comentarios de los afectados por las decisiones de curación y moderación de contenido es de vital importancia para una correcta evaluación de los riesgos para los derechos humanos.

### Qué proponemos

Para facilitarlo, las plataformas deberían:

- ▶ Habilitar la participación significativa de los usuarios de manera oportuna y en diferentes etapas de la creación, implementación y evaluación de las normas y desarrollos tecnológicos relativos a la gobernanza de contenido
- ▶ Hacer que diferentes grupos de usuarios se involucren, particularmente los más afectados por ciertas normas, decisiones o tecnologías
- ▶ Designar puntos de contacto locales para recibir comentarios y responder a los usuarios y a la sociedad civil, cuando sea posible. Esos puntos de contacto deberían manejar el idioma local y tener completo conocimiento sobre la realidad social y cultural de la región
- ▶ Formar parte de iniciativas realmente participativas, abiertas, independientes y transparentes, dedicadas a aumentar la transparencia y la supervisión en relación con las decisiones de moderación y curación de contenido

43

## C. Recomendaciones para una corregulación respetuosa de los derechos humanos

Los mecanismos de corregulación cumplen una función importante en la gobernanza de contenido, tanto los formales como los informales. Esto resulta evidente en los casos en que se necesita la cooperación entre los gobiernos y las plataformas para abordar algunos de los desafíos más complejos en la regulación: la necesidad de tomar acciones rápidamente y de coordinar de manera global para lidiar con el contenido ilegal, entre otras cuestiones.

Sin embargo, para que la corregulación sea una herramienta que aborde las necesidades sociales y extienda los derechos de los usuarios, se deben tener en cuenta algunas consideraciones. Los actores involucrados en los esfuerzos de corregulación deben cumplir con las obligaciones y responsabilidades de derechos humanos concernientes a sus funciones, como se explicó en secciones anteriores de este documento. Asimismo, se deben abordar algunas consideraciones especiales.

### 1. ADOPTAR MARCOS LEGALES PARTICIPATIVOS, CLAROS Y TRANSPARENTES

Al igual que en las recomendaciones para los estados, todo modelo corregulatorio debe tener sus bases en un marco legal vinculante adoptado

por los actores estatales. Esto daría paso a los mecanismos de rendición de cuentas necesarios que evitan que los actores privados tomen decisiones poco transparentes y posiblemente arbitrarias.

Un marco legal apropiado debería establecer salvaguardas para los usuarios y un mecanismo de supervisión independiente para todo modelo corregulatorio. Se deben establecer y distinguir claramente las obligaciones de los estados y las responsabilidades de los intermediarios para proteger los derechos humanos de los usuarios.

Todos los enfoques de corregulación deberían respetar los derechos humanos internacionales, los requisitos de transparencia apropiados, y el principio de participación significativa.

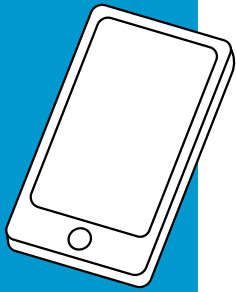
## **2. NO CAMBIAR NI TERGIVERSAR LAS RESPONSABILIDADES DE LOS ACTORES**

Co-regulatory initiatives should never be understood as a replacement for the duties or responsibilities of the individual actors involved. Governments should stop allowing or encouraging private actors to decide upon the legality of user-generated content and its restriction.

In all cases, actors should enable users' access to effective remedy, including judicial remedy and redress.

## **3. PREVENIR ABUSOS**

Los estados deben evitar cualquier acción que pueda resultar en un abuso de las medidas corregulatorias, como el monitoreo general del contenido, el cumplimiento excesivo por parte de los intermediarios, y la remoción excesiva del contenido generado por los usuarios.



## VI. Conclusión

El crecimiento del uso y la facilidad de acceso a los servicios de Internet ha traído consigo nuevos desafíos para el espacio público en línea. La idea de un espacio abierto y sin fronteras donde las personas pueden gozar de la libertad de acceso a la información y expresar sus opiniones ha entrado en conflicto con las amenazas a esos ideales, que surgen de las acciones de los guardianes de la información y la explotación de la tecnología con fines violentos y dañinos.

Resulta evidente que todos los actores en el espacio de la información deben tomar medidas para lidiar con los efectos negativos de algunos usos de Internet. Esto incluye a los gobiernos y las plataformas que actúan como intermediarios de la información en línea, pero también a otros actores que contribuyen a la formación del discurso público, como los medios tradicionales y los actores políticos.

Celebramos la predisposición de algunos gobiernos y plataformas de revisar sus deberes y responsabilidades y de facilitar un diálogo participativo para encontrar soluciones al contenido en línea ilegal, dañino y violento. No obstante, también nos preocupan los enfoques apresurados, las iniciativas poco transparentes, y las soluciones propuestas que entran en conflicto con derechos humanos básicos.

Lamentablemente, no existe una panacea para resolver problemas complejos como la proliferación de contenido ilegal, violento o dañino en Internet. Las soluciones que se basan únicamente en la censura o la criminalización de la actividad de los usuarios pueden ser contraproducentes y afectar los derechos de grupos vulnerables, medios independientes, artistas, activistas, y defensores de derechos humanos, entre otros. Proteger y preservar sus opiniones es tan importante para el fortalecimiento de nuestras democracias como lo es la seguridad de las personas y la lucha contra el comportamiento delictivo.

Los gobiernos deben escuchar atentamente y actuar de manera democrática y transparente. Las propuestas regulatorias deben estar basadas en evidencia e ir más allá de la criminalización, considerando también la concentración económica, los modelos de negocio, la protección de datos, la educación de los usuarios, y la inclusión social como factores clave en el abordaje de la ilegalidad y la violencia en línea. La actividad dañina en Internet es, a menudo, un reflejo de problemas sociales. Los actores políticos pueden hacer más para lidiar con estos problemas abandonando la retórica incendiaria que lleva a la división y potencialmente a la incitación a la violencia.

Es necesario que las plataformas asuman y ajusten sus responsabilidades de acuerdo con el impacto que tienen en el discurso público. Esto incluye la observancia de los estándares de derechos humanos en el contexto de sus prácticas de curación y moderación de contenido e incorporar transparencia, privacidad, protección de datos y evaluación de impactos en todos sus productos y prácticas. Otras plataformas que participan y habilitan el discurso público, como los medios tradicionales, también deben asumir el compromiso de mejorar sus prácticas.

Todas las medidas tomadas por los gobiernos y las plataformas respecto de la gobernanza de contenido deberían tener un objetivo específico en mente: facilitar un espacio en línea que sea diverso, libre, abierto y seguro. Esperamos que esta guía pueda ayudar a garantizar un discurso público sano, fortalecer las democracias, y preservar el goce de los derechos humanos mediante la libre expresión y el acceso a la información para todos.



## VII. Glosario

Con el objetivo de ayudar a los tomadores de decisiones en el debate sobre enfoques de gobernanza de contenido basada en los derechos humanos, presentamos a continuación un breve glosario de términos que utilizamos en este documento como nosotros los definimos.

**Autorregulación:** una forma de gobernanza de contenido ejercida por las plataformas en línea en base a sus propios términos de servicio o "pautas comunitarias", usualmente mediante prácticas de moderación de contenido que hacen uso tanto de herramientas automatizadas como de moderadores humanos.

**Conocimiento efectivo:** se refiere al momento en que una plataforma en línea toma conciencia de la existencia de contenido ilegal específico en la plataforma. El conocimiento efectivo usualmente se obtiene mediante notificaciones que provienen de diversas fuentes, incluidas las órdenes judiciales o los alertadores de confianza. En este documento, especificamos que únicamente una notificación válida constituye la toma de conocimiento efectivo.

**Contenido generado por los usuarios de ilegalidad manifiesta:** decimos que el contenido es de ilegalidad manifiesta cuando es reconocible como tal sin un mayor análisis legislativo o factual por parte de las plataformas. Incluye, por ejemplo, el material que muestra abuso de menores.

**Corregulación:** los regímenes de corregulación se pueden definir como una autorregulación que las autoridades públicas promueven, apoyan y, en ocasiones, monitorean activamente. Algunos ejemplos típicos de los mecanismos de corregulación son los códigos de conducta voluntarios que resultan del diálogo entre actores privados y autoridades estatales nacionales o regionales.

**Curación de contenido:** una forma de autorregulación. Con la curación de contenido, las plataformas en línea toman decisiones respecto del alcance, la prominencia o la amplificación de cierto material. Así, se determina cuáles y cuántos usuarios están expuestos al material selecto y de qué forma se presenta. La curación de contenido requiere el uso de datos sobre las interacciones de los usuarios en la plataforma y, a menudo, es llevada a cabo por sistemas de machine learning.

**Gobernanza de contenido:** en este documento, la "gobernanza" de contenido se refiere a interacciones y procesos complejos delineados por las partes interesadas públicas y privadas para crear y aplicar normas para regular el contenido en línea. Usamos esta unidad terminológica como término general que abarca tres enfoques regulatorios específicos sobre el contenido generado por los usuarios: la autorregulación, la corregulación y la regulación estatal.

**Notificación válida:** para que una notificación sobre contenido sea válida, debe contener suficiente información para que las plataformas actúen. Debe ser precisa y estar justificada adecuadamente. Las condiciones que necesita una notificación para ser válida deben estar especificadas en la ley.

**Notificación y acción:** término general que hace referencia a una variedad de

mecanismos que tienen como objetivo eliminar contenido ilegal, infractor o potencialmente dañino en Internet. Los mecanismos de notificación y acción permiten a las plataformas en línea tomar decisiones sobre qué contenido seguirá siendo accesible y qué contenido deberá ser eliminado sin implicar una supervisión pública.

La notificación puede ser presentada directamente por un usuario, por ejemplo, para alertar sobre contenido que puede violar los términos de servicio, o por una autoridad pública mediante una orden judicial, tras la cual la plataforma debe tomar medidas.

**Plataformas en línea:** permiten a las personas usar tecnologías de información y comunicación para facilitar diversos tipos de interacciones entre sus usuarios.

Existen en una variedad de formas, desde redes sociales hasta plataformas de colaboración económica. Las interacciones entre los usuarios generan datos que luego las plataformas recopilan y usan. Si bien el uso de esos datos mejora significativamente la experiencia del usuario, también plantea una gran cantidad de preocupaciones de derechos humanos, que mencionamos anteriormente en este documento.

**Principio de certeza legal:** garantiza que la aplicación de la ley en una situación específica sea predecible y, por lo tanto, debe brindar a las partes claridad sobre cómo regular su propia conducta. La certeza legal es un principio fundamental recogido en las constituciones nacionales y también en el derecho internacional. Es una precondition esencial para el correcto funcionamiento del estado de derecho en una sociedad democrática.

**Principio de proporcionalidad:** regula la manera en que los estados ejercen sus poderes. Bajo este principio, todas las acciones de las autoridades públicas que puedan limitar el ejercicio de un derecho humano deben cumplir los siguientes requisitos: ser necesarias en una sociedad democrática y ser las menos restrictivas hacia los derechos humanos de los usuarios; perseguir un objetivo legítimo establecido en las leyes de derechos humanos; lograr un equilibrio entre los medios usados y los objetivos que se quieren alcanzar.

**Prohibición de monitoreo general:** Los estados no deberían imponer a las plataformas en línea la obligación de monitorear todo el contenido generado por los usuarios que estas transmiten o almacenan. La prohibición de monitoreo general existe actualmente en virtud del marco legal europeo.

**Proveedor de contenido:** el ente que sube el contenido a la plataforma, ya sea un individuo, una empresa u otra entidad legal.

**Puerto seguro:** una disposición en las leyes de responsabilidad legal de los intermediarios que protege a las plataformas en línea de ser responsabilizadas legalmente de la actividad ilegal o potencialmente dañina de los usuarios mediante sus servicios si dichas plataformas cumplen ciertos criterios (lo que debería estar prescripto por ley).

**Regulación estatal:** la regulación estatal hace referencia a cualquier instrumento jurídicamente vinculante o regulatorio que promulgado por instituciones públicas locales, nacionales o regionales mediante sus procesos legislativos o regulatorios.



**Responsabilidad legal de los intermediarios:** generalmente, se refiere a la responsabilidad jurídica de las plataformas en línea respecto de la actividad ilegal o potencialmente dañina de los usuarios a través de su servicio. Los regímenes de responsabilidad legal comenzaron a incorporarse en legislaciones nacionales en la década de 1990, y la definición de este término puede diferir en las distintas jurisdicciones alrededor del mundo. Este documento se refiere específicamente a los modelos de responsabilidad legal de Estados Unidos y la Unión Europea, debido a la influencia mundial que han tenido en la formulación de respuestas legales a la gobernanza de contenido.

**Términos de servicio:** un conjunto de normas en forma de acuerdo legal entre la plataforma en línea y el usuario que desea usar sus servicios.

## Notas finales

1. Barlow, J.P. (1996). A Declaration of the Independence of Cyberspace. Disponible en <https://www.eff.org/cyberspace-independence>
2. Gillespie, T. (2018). Custodians of the internet, p. 30-31. Disponible en [https://www.researchgate.net/publication/327186182\\_Custodians\\_of\\_the\\_internet\\_Platforms\\_content\\_moderation\\_and\\_the\\_hidden\\_decisions\\_that\\_shape\\_social\\_media](https://www.researchgate.net/publication/327186182_Custodians_of_the_internet_Platforms_content_moderation_and_the_hidden_decisions_that_shape_social_media)
3. Este es un concepto amplio que abarca diferentes tipos de plataformas que actúan a distintos niveles de la pila de comunicaciones en línea. Los proveedores de aplicaciones (como las redes sociales o los servicios de búsqueda), las páginas web, los foros, los proveedores de conexión a Internet, los proveedores de servicio de seguridad de Internet e infraestructura, e incluso los fabricantes de dispositivos pueden considerarse intermediarios en las comunicaciones de los usuarios. Cabe recordar que no todos intervienen de manera activa en el contenido de los intercambios de información entre los usuarios. A los efectos de este documento, usamos el término "plataforma en línea" para hablar de plataformas y servicios que actúan como intermediarios del discurso generado por los usuarios y que toman decisiones de manera activa acerca del tipo o la prominencia del contenido que permiten en sus servicios.
4. Como los derechos económicos, sociales y culturales. Consulte Lara, J.C. (2015). Internet access and economic, social, and cultural rights. Association for Progressive Communications (APC). Partes 3 y 4. Disponible en [https://www.apc.org/sites/default/files/APC\\_ESCR\\_Access\\_Juan%20Carlos%20Lara\\_September2015%20\(1\).pdf](https://www.apc.org/sites/default/files/APC_ESCR_Access_Juan%20Carlos%20Lara_September2015%20(1).pdf)
5. El Libro blanco sobre daños en línea de 2019 del Reino Unido emplea el término "daños en línea", que incluye contenido ilegal, pero también cubre los comportamientos de los usuarios que se estiman dañinos, pero no necesariamente ilegales. Consulte la Secretaría de Estado de Tecnología, Cultura, Medios y Deportes del Reino Unido y otros (2019). Online Harms White Paper. Disponible en [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/793360/Online\\_Harms\\_White\\_Paper.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf)
6. Facebook. (n.d.). Community Standards. Part III, Objectionable Content. Retrieved from [https://www.facebook.com/communitystandards/objectionable\\_content](https://www.facebook.com/communitystandards/objectionable_content)
7. Ranking Digital Rights. (2019). Corporate Accountability Index. Chapter 4 Freedom of Expression. Retrieved from <https://rankingdigitalrights.org/index2019/report/freedom-of-expression/>
8. La Línea de Ayuda de Seguridad Digital es un servicio provisto por Access Now que ofrece asistencia técnica directa y en tiempo real y asesoramiento a grupos de sociedad civil y activistas, organizaciones de medios, periodistas, blogueros, y defensores de derechos humanos. La Línea de Ayuda brinda asistencia para que las organizaciones evalúen los riesgos de seguridad digital, resuelvan problemas existentes y adopten mejores prácticas, ayudándolas a desarrollar una mentalidad segura para el futuro. El servicio está disponible las 24 horas del día en nueve idiomas: inglés, español, francés, alemán, portugués, ruso, tagalo, árabe e italiano. Para obtener más información sobre la Línea de Ayuda de Seguridad Digital de Access Now, visit <https://www.accessnow.org/help/>
9. "La gobernanza se puede referir, de manera abstracta, a todos los procesos de regulación. Brinda un enfoque sobre las instituciones formales de gobierno con reconocimiento de actividades más diversas que borran los límites entre los estados y la sociedad. Centra la atención en los complejos procesos e interacciones que tienen que ver con la regulación. Más concretamente, la gobernanza también puede referirse al surgimiento de nuevos procesos de regulación que son híbridos y multijurisdiccionales, con múltiples partes interesadas que trabajan en conjunto en redes. Describe los cambios recientes en el mundo" Bevir, Mark (2012). Governance: A Very Short Introduction. OUP Oxford.
10. Estas normas pueden tener distintos nombres o hacer referencia a otros documentos como "pautas comunitarias" que establecen las normas para la expresión aceptable en la plataforma.
11. En este documento, el concepto de discurso incluye la expresión del usuario en todas sus formas técnicamente posibles en una plataforma, como textos, imágenes, videos, etc.
12. Este tipo de decisión también se conoce como "clasificación de contenido" o "distribución de contenido", entre otras denominaciones.
13. Las plataformas pueden tomar decisiones sobre la exposición o el alcance en base al aumento en la "relevancia", reaccionando a los intereses de los usuarios, decisiones de políticas públicas, etc. Con respecto a los métodos de curación de contenido, las plataformas podrían determinar qué priorizar independientemente o implementar un sistema de toma de decisiones automatizado. Podrían priorizar cierto contenido para una categoría de usuarios para ajustarlo a las preferencias de cada individuo. En todos los casos, el desconocimiento de los usuarios respecto del sesgo algorítmico representa una preocupación. Consulte Bozdag,

E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3), 209–227.

14. Mardsen, C. (2011). *Internet co-regulation: European law, regulatory governance, and legitimacy in cyberspace*, p. 44. Cambridge University Press.
15. Christchurch Call to eliminate terrorist and violent extremist content online (2019). Disponible en <https://www.christchurchcall.com/>
16. Global Internet Forum to Counter Terrorism (n.d.). Retrieved from <https://www.gifct.org/>
17. Consejo de Derechos Humanos (2018). Informe de la Relatora Especial sobre la violencia contra la mujer, sus causas y consecuencias acerca de la violencia en línea contra las mujeres y las niñas desde la perspectiva de los derechos humanos. A/HRC/38/47. Retrieved from [https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session38/Documents/A\\_HRC\\_38\\_47\\_EN.docx](https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session38/Documents/A_HRC_38_47_EN.docx)
18. Incluyendo los casos de censura y acoso que involucran a YouTube, Facebook e Instagram.
19. Consejo de Derechos Humanos (2018). Op.cit.
20. Karp, P. (2019). Australia passes social media law penalizing platforms for violent content. *The Guardian*. Disponible en <https://www.theguardian.com/media/2019/apr/04/australia-passes-social-media-law-penalising-platforms-for-violent-content>
21. Consulte, por ejemplo, la creación de sistemas de referencia en la Regulación propuesta por la UE para la lucha contra la diseminación del contenido terrorista en línea.
22. Tribunal Europeo de Derechos Humanos (2018). *Stomakhin v. Russia*, No. 52273/07, párrs. 93-131; (2008). *Leroy v. France*, No. 36109/03; (2014). *M'Bala M'Bala v. France*, No. 25239/13, párrs. 37-39.
23. Parlamento alemán (2017). *Netzdurchsetzungsgesetz -Ley de Aplicación de Redes-*. Disponible en el Archivo de Derecho Alemán en <https://germanlawarchive.iuscomp.org/?p=1245>
24. Bychawska-Siniarska, D. (2017). Protecting the Right to Freedom of Expression under the European Convention on Human Rights. 4.5. Disponible en <https://rm.coe.int/handbook-freedom-of-expression-eng/1680732814>
25. Pallero, J. (2018). Honduras: new bill threatens to curb online speech. *Access Now*. Disponible en <https://www.accessnow.org/honduras-new-bill-threatens-curb-online-speech/>
26. Taye, B. & Jit Singh Chima, R. (2018). Bangladesh wants to fight “fake news” during elections with internet shutdowns. We can’t let that happen. *Access Now*. Disponible en <https://www.accessnow.org/bangladesh-wants-to-fight-fake-news-during-elections-with-internet-shutdowns-we-cant-let-that-happen/>
27. Douek, E. (2017). Germany’s Bold Gambit to Prevent Online Hate Crimes and Fake News Takes Effect. *Lawfare*. Disponible en <https://www.lawfareblog.com/germanys-bold-gambit-prevent-online-hate-crimes-and-fake-news-takes-effect>
28. Kretschmer, M. & Erickson, K. (2018). How much do we know about notice-and-takedown? New study tracks YouTube removals. *Kluwer Copyright Blog*. Retrieved from <http://copyrightblog.kluweriplaw.com/2018/06/12/much-know-notice-takedown-new-study-tracks-youtube-removals/>
29. Newton, C. (2019). The Trauma Floor. *The Verge*. Disponible en <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>
30. Este es el caso, por ejemplo, en las bajas automatizadas de discurso político y voces marginalizadas en base a filtros de carga de derecho de autor. Consulte Reda, J. (2017). When filters fail: These cases show we can’t trust algorithms to clean up the internet. Disponible en <https://juliareda.eu/2017/09/when-filters-fail/>
31. Tribunal Europeo de Derechos Humanos (2016). *Cengiz and Others v. Turkey*, párr. 49. Disponible en <https://hudoc.echr.coe.int/eng?i=001-159188>
32. Corte Suprema de los Estados Unidos (2017). *Packingham v. North Carolina*. Disponible en [https://www.supremecourt.gov/opinions/16pdf/15-1194\\_08l1.pdf](https://www.supremecourt.gov/opinions/16pdf/15-1194_08l1.pdf)
33. Consulte Roger Denson, G. (2017). *Courbet’s Origin Of The World Still Too Scandalous For Media-Savvy Facebook*. *Huffington Post*. Disponible en [https://www.huffpost.com/entry/courbets-1866-the-origin\\_b\\_1087604](https://www.huffpost.com/entry/courbets-1866-the-origin_b_1087604) y Levin, S. et al. (2016). Facebook backs down from ‘napalm girl’ censorship and reinstates photo. *The Guardian*. Disponible en <https://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo>
34. Consejo de Derechos Humanos (2018). Op.cit.
35. Wong, C. Solon, O. (2018). Facebook releases content moderation guidelines – rules long kept secret. *The Guardian*. Disponible en <https://www.theguardian.com/technology/2018/apr/24/>

- facebook-releases-content-moderation-guidelines-secret-rules
36. Organización de los Estados Americanos (2019). Comunicado de prensa R116/19. La Oficina del Relator Especial repudia el cierre de Radio Caracas 750 AM, la censura de canales de televisión, las restricciones en Internet, y la detención de periodistas en Venezuela. Disponible en <https://www.oas.org/en/iachr/expression/showarticle.asp?artID=1140&IID=1>
  37. Wu, T. (2017). The Attention Economy and the Law. Retrieved from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2941094](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2941094)
  38. Asamblea General de las Naciones Unidas (2018). Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y expresión. A/73/348. Pág. 11. Disponible en <https://freedex.org/wp-content/blogs.dir/2015/files/2018/10/AI-and-FOE-GA.pdf>
  39. Fisher, M. Taub, A. (2019). How YouTube Radicalized Brazil. The New York Times. Con citas del trabajo de Jonas Kaiser y Yasodara Córdova, con Adrian Rauchfleisch de la Universidad Nacional de Taiwán. Disponible en <https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html> Además, consulte "On YouTube's Digital Playground, an Open Gate for Pedophiles" de los mismos autores, disponible en <https://www.nytimes.com/2019/06/03/world/americas/youtube-pedophiles.htm>
  40. Goel, V. (2014). Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry. The New York Times. Con citas de la investigación de Adam D. I. Kramer et. al. Disponible en <https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html>
  41. Goodman, R. (2017). Facebook's ad-targeting problems prove how easy it is to discriminate online. NBC Think. Disponible en <https://www.nbcnews.com/think/opinion/facebook-s-ad-targeting-problems-prove-how-easy-it-discriminate-ncna825196>
  42. Matsa, K. Shearer, E. (2018). News Use Across Social Media Platforms 2018. Pew Research Center. Retrieved from <https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/>
  43. Access Now, Civil Liberties Union for Europe, European Digital Rights (2018). Informing the "Disinformation" Debate. Disponible en [https://edri.org/files/online\\_disinformation.pdf](https://edri.org/files/online_disinformation.pdf)
  44. omisión Europea (2016). Código de Conducta para la Lucha contra la Incitación Ilegal al Odio en Internet. Disponible en [https://ec.europa.eu/newsroom/just/document.cfm?doc\\_id=42985](https://ec.europa.eu/newsroom/just/document.cfm?doc_id=42985). El Código de Conducta exige que las plataformas en línea revisen todas las notificaciones válidas presentadas que denuncien discurso de odio ilegal en línea en menos de 24 horas y eliminen o inhabiliten el acceso a dicho contenido.
  45. Pallero, J. (2019). Access Now on the ChristChurch Call: rights, wrongs, and what's next. Disponible en <https://www.accessnow.org/access-now-on-the-christchurch-call-rights-wrongs-and-whats-next/>
  46. Diaz, A. (2019). Global Internet Forum to Counter Terrorism's 'Transparency Report' Raises More Questions Than Answers. Just Security. Disponible en <https://www.justsecurity.org/66298/gifct-transparency-report-raises-more-questions-than-answers/>
  47. Comisión Europea (2018). Recomendación de la Comisión sobre las medidas destinadas a luchar, de manera eficaz, contra el discurso de odio ilegal en línea. Disponible en <https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-measures-effectively-tackle-illegal-content-online>
  48. Comisión Europea (2019). Combatir el terrorismo en línea: el Foro de la UE sobre Internet se compromete a aplicar un Protocolo de crisis a escala de la UE. Disponible en [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_19\\_6009](https://ec.europa.eu/commission/presscorner/detail/en/ip_19_6009)
  49. Comisión Europea (2019). Combatir el terrorismo en línea: el Foro de la UE sobre Internet se compromete a aplicar un Protocolo de crisis a escala de la UE. Disponible en [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_19\\_6009](https://ec.europa.eu/commission/presscorner/detail/en/ip_19_6009)
  50. La prisa por cumplir con las presiones políticas y otras necesidades asociadas con el contenido ilegal o dañino ha hecho que las empresas adopten prácticas que carecen de transparencia y proporcionalidad. La remoción de evidencia sobre violaciones de derechos humanos como parte del esfuerzo por frenar el contenido terrorista es un ejemplo de esto. Consulte York, J., Al Jaloud, A., Al Kathib, H., Kayyali, D. (2019). Caught in the Net: The Impact of "Extremist" Speech Regulations on Human Rights Content. Disponible en <https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content#Blunt>
  51. Feld, H. (2018). Part III: Cost of Exclusion as a Proxy for Dominance in Digital Platform Regulation. Public Knowledge. Retrieved from <https://www.publicknowledge.org/news-blog/blogs/part-iii-cost-of-exclusion-as-a-proxy-for-dominance-in-digital-platform-reg>

52. Por ejemplo, la Directiva sobre Comercio Electrónico estipula en el Considerando 42 que la actividad del intermediario debe ser de una naturaleza meramente técnica, automática y pasiva. El intermediario no puede tener conocimiento ni control de la información transmitida y almacenada. El Tribunal de Justicia de la Unión Europea (TJUE) aclaró la distinción entre un intermediario activo y uno pasivo: un intermediario pasivo no tiene conocimiento ni control sobre la información que transmite o almacena. Consulte los casos C-236/08 y C-237/08 *Google France v Louis Vuitton et. al.* (2010) ECR I-241; Caso C-324/09 *L'Oréal v eBay International* (2011) ECR I-6011; Caso C-291/13 *Papasavas* (2014) EUECJ C-291/13
53. Gillespie, T. Op.cit.
54. Barzilai-Nahon, K. (2008). Towards a Theory of Network Gatekeeping: A Framework for Exploring Information Control. *Journal of the American Society for Information Science and Technology*, Vol. 59(9), p. 1493–1512. Laidlaw, E. (2015). A Framework for Identifying Internet Information Gatekeeper. *International Review of Law, Computers & Technology*, Vol. 24, No. 3, 2010. Retrieved from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2667902](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2667902)
55. Prince, M. (2019). Terminating Service for 8Chan. Cloudflare. Retrieved from <https://blog.cloudflare.com/terminating-service-for-8chan>
56. "La legislación aplicable a los intermediarios de Internet y a sus relaciones con los estados y los usuarios debe ser accesible y previsible". Consejo de Europa (2018), op. cit.; "Los gobiernos deben publicar en Internet toda legislación, decisiones de políticas y cualquier otro tipo de regulación que sea relevante respecto de la responsabilidad de los intermediarios en forma oportuna y en formatos accesibles". Principios de Manila sobre responsabilidad de los intermediarios (2015). Disponible en <https://www.manilaprinciples.org/>
57. "La legislación aplicable a los intermediarios de Internet y a sus relaciones con los estados y los usuarios debe ser accesible y previsible". Consejo de Europa (2018), op. cit.; "Los gobiernos deben publicar en Internet toda legislación, decisiones de políticas y cualquier otro tipo de regulación que sea relevante respecto de la responsabilidad de los intermediarios en forma oportuna y en formatos accesibles". Principios de Manila sobre responsabilidad de los intermediarios (2015). Disponible en <https://www.manilaprinciples.org/>
58. Cortes de Distrito de los Estados Unidos, S.D. Nueva York (1991). *Cubby, Inc. v. CompuServe Inc.*; Corte Suprema, Condado de Nassau, Nueva York (1995). *Stratton Oakmont, Inc. v. Prodigy Servs. Co.* Disponible en <https://h2o.law.harvard.edu/cases/4540>
59. Corte de Apelaciones de Estados Unidos, Primer Circuito (2016). *Jane Doe No. 1 v. Backpage.com, LLC.* Disponible en <https://caselaw.findlaw.com/us-1st-circuit/1728752.html>. En este caso, el Primer Circuito resolvió que la Sección 230 protege las elecciones de los sitios web como hablantes y medios de publicación, indicando lo siguiente: "El Congreso fue claro al promulgar la CDA y elegir otorgar protecciones amplias a los medios de publicación en Internet. Demostrar que un sitio web opera mediante un modelo de negocio rimbombante no es suficiente para quitar esas protecciones".
60. Tribunal de Justicia de la Unión Europea (2012). *SABAM v. Netlog.* Caso C-360/10. Disponible en <http://curia.europa.eu/juris/document/document.jsf?docid=119512&doclang=EN>
61. Principios de Manila (2015), op.cit.
62. Consejo de Europa (2018). Recomendación CM/Rec(2018)2 del Comité de Ministros a los Estados miembros sobre las funciones y las responsabilidades de los intermediarios de Internet. Disponible en <https://rm.coe.int/1680790e14intermediaries>.
63. Consejo de Derechos Humanos (2018). Op.cit.
64. Tribunal de Justicia de la Unión Europea (2011). *L'Oréal SA v. eBay International AG*, C-324/09. Disponible en <http://curia.europa.eu/juris/liste.jsf?num=C-324/09>
65. Court of Justice of the European Union, op. cit.
66. Tribunal de Justicia de la Unión Europea (2016). *McFadden v. Sony Music*, Caso C-484/14. Disponible en <http://curia.europa.eu/juris/document/document.jsf?docid=183363&doclang=EN&mode=lst&occ=first>
67. Kuczerawy, A. (2019). Intermediary Liability and Freedom of Expression in the EU: Concepts and Safeguards, pp. 296-297. Disponible en <https://intersentia.com/en/intermediary-liability-and-freedom-of-expression-in-the-eu-from-concepts-to-safeguards.html>
68. Kuczerawy, A. (2019), op. cit., p. 305.
69. Pacto Internacional de Derechos Civiles y Políticos (16 de diciembre de 1966). Disponible en <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx>
70. Convenio Europeo para la Protección de los Derechos Humanos (4 de noviembre de 1950). Disponible en <http://www.hri.org/docs/ECHR50.html>
71. Comisión Europea. (2012). Un marco coherente para aumentar la confianza en el mercado único digital del comercio electrónico y los servicios en línea, p.13, nota al pie 49. Disponible en <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52011DC0942>

72. Kuczerawy, A. (2019). Intermediary Liability and Freedom of Expression in the EU: Concepts and Safeguards, p. 38. Volume 3 of Ku Leuven Centre for It & IP Law. Intersentia.
73. Aleksandra Kuczerawy, Christina Angelopoulos, Stijn Smets, y Joris van Hoboken, entre otros.
74. Este mecanismo para las violaciones de derechos de autor fue adoptado por Canadá en la disposición central de la reforma de derecho de autor digital de 2012. Ley de Modernización de los Derechos de Autor (2012). Disponible en [https://laws-lois.justice.gc.ca/eng/annualstatutes/2012\\_20/fulltext.html](https://laws-lois.justice.gc.ca/eng/annualstatutes/2012_20/fulltext.html)
75. Organization for Economic Cooperation and Development. (2011). The Role of Internet Intermediaries in Advancing Public Policy Objectives – Forging partnerships for advancing policy objectives for the Internet economy, p. 57. Retrieved from <https://www.oecd.org/sti/ieconomy/theroleofinternetintermediariesinadvancingpublicpolicyobjectives.htm>
76. For the overview of national implementations of notice-and-notice procedure, please consult Kuczerawy, A., op.cit.
77. Angelopoulos, C., Smet, S. (2016). Notice-and-Fair-Balance: How to Reach a Compromise between Fundamental Rights in European Intermediary Liability, p. 23. Disponible en [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2944917](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2944917)
- 78 Principios de Manila. (2015), op.cit., Comisión Europea (2018). Recomendación de la Comisión sobre las medidas destinadas a luchar, de manera eficaz, contra el discurso de odio ilegal en línea. Disponible en <https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-measures-effectively-tackle-illegal-content-online>
79. Comisión Europea. (2017). Comunicación sobre la lucha contra el contenido ilegal en Internet – Hacia una mejor responsabilidad de las plataformas en línea, p. 8. Disponible en <https://ec.europa.eu/digital-single-market/en/news/communication-tackling-illegal-content-online-towards-enhanced-responsibility-online-platforms>
80. Kuczerawy, A. (2019), op. cit., p. 307.
81. Consejo de la Unión Europea. (2018). Reglamento del Parlamento Europeo y del Consejo para la prevención de la difusión de contenidos terroristas en línea
82. Consejo de la Unión Europea. (2018). Op. cit.
83. Consejo de la Unión Europea. (2016). Comparative study on blocking, filtering, and take-down of illegal internet content, p. 16. Retrieved from <https://edoc.coe.int/en/internet/7289-pdf-comparative-study-on-blocking-filtering-and-take-down-of-illegal-internet-content-.html>
84. Consejo de la Unión Europea. (2018). Op. cit.
85. Wagner, B. et al. (2020). Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act. FAT\* '20, January 27–30, 2020, Barcelona, Spain.
86. Ranking Digital Rights. (2017). Submission to UN Special Rapporteur for Freedom of Expression and Opinion David Kaye: Content Regulation in the Digital Age. Disponible en <https://rankingdigitalrights.org/wp-content/uploads/2018/01/RDR-2018-David-Kaye-Submission.pdf>
87. Comisión Europea. (2017). Comunicación sobre la lucha contra el contenido ilegal en Internet – Hacia una mejor responsabilidad de las plataformas en línea. Disponible en <https://ec.europa.eu/digital-single-market/en/news/communication-tackling-illegal-content-online-towards-enhanced-responsibility-online-platforms>
88. Estos principios se ajustan considerablemente a aquellos recomendados por el Relator Especial de la ONU sobre la promoción y la protección del derecho a la libertad de opinión y expresión y aquellos recogidos en los Principios de Santa Clara sobre la transparencia y la rendición de cuentas en la moderación de contenido (s.f.). Disponible en <https://santaclaraprinciples.org/>
89. Mirani, L. (2015). Millions of Facebook users have no idea they're using the internet. Quartz. Disponible en <https://qz.com/333313/millions-of-facebook-users-have-no-idea-theyre-using-the-internet/>
90. Consejo de Derechos Humanos. (2008). Proteger, respetar, remediar: un marco para las empresas y los derechos humanos. Informe del Representante Especial del Secretario General para la cuestión de los derechos humanos y las empresas transnacionales y otras empresas, John Ruggie. A/HRC/8/5. Disponible en <https://www2.ohchr.org/english/bodies/hrcouncil/docs/11session/A.HRC.11.13.pdf>. Informe del Representante Especial del Secretario General para la cuestión de los derechos humanos y las empresas transnacionales y otras empresas, John Ruggie. A/HRC/8/5. Disponible en <https://www2.ohchr.org/english/bodies/hrcouncil/docs/11session/A.HRC.11.13.pdf>
91. Consejo de Derechos Humanos. (2009). Empresas y los Derechos Humanos: Hacia la operativización del marco "Proteger, respetar y remediar" – Informe del Representante Especial del Secretario General para la cuestión de los derechos humanos y las empresas transnacionales y otras empresas. A/HRC/11/13. Disponible en <https://www2.ohchr.org/english/bodies/hrcouncil/docs/11session/A.HRC.11.13.pdf>
92. Consejo de la Unión Europea. Op. cit.



Access Now (<https://www.accessnow.org>) defiende y extiende los derechos digitales de los usuarios en riesgo alrededor del mundo. Mediante la combinación de apoyo técnico directo, campañas globales, el análisis integral de políticas públicas, el financiamiento a grupos locales emergentes y eventos como RightsCon, luchamos por los derechos humanos en la era digital.