Naciones Unidas A/HRC/38/35

Asamblea General

Distr. general
6 de abril de 2018
Español
Original: inglés

Consejo de Derechos Humanos
38º período de sesiones
18 de junio a 6 de julio de 2018
Tema 3 de la agenda
Promoción y protección de todos los derechos humanos, civiles, políticos, económicos, sociales y culturales, incluido el derecho al desarrollo

Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión

Nota de la Secretaría

La Secretaría tiene el honor de transmitir al Consejo de Derechos Humanos el informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, David Kaye, de conformidad con la resolución 34/18 del Consejo. En su informe, el Relator Especial aborda la cuestión de la regulación del contenido en línea generado por los usuarios. El Relator Especial recomienda que los Estados garanticen un entorno propicio para la libertad de expresión en línea y que las empresas apliquen las normas de derechos humanos en todas las etapas de sus operaciones. El derecho de los derechos humanos ofrece a las empresas los instrumentos necesarios para que puedan expresar sus posiciones de manera que se respeten los principios democráticos y oponerse a las exigencias autoritarias. Como mínimo, las empresas y los Estados deberían esforzarse por conseguir una transparencia mejorada de forma radical, desde el establecimiento de las normas hasta su aplicación, a fin de garantizar la autonomía de los usuarios a medida que las personas ejercen cada vez con mayor frecuencia sus derechos fundamentales en línea.

GE.18-05436 (S) 200418 250418
1805436

Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión

Índice

I. Introducción

1. En los albores de la era digital, John Perry Barlow declaró que Internet traería consigo “un mundo en el que cualquier persona, en cualquier lugar podrá expresar sus creencias, sin importar cuán singulares sean, sin temor a ser obligada a guardar silencio o a expresar conformidad”[1]. Aunque Internet sigue siendo el instrumento más importante de la historia en lo que se refiere al acceso a la información a nivel mundial, hoy día es difícil abrazar una creencia tan profunda sobre el mundo en línea. El público percibe odio, abuso y desinformación en el contenido generado por los usuarios. Los Gobiernos detectan el reclutamiento de terroristas o la realización de incómodas actividades de disidencia y oposición. Las organizaciones de la sociedad civil observan cómo se subcontratan funciones públicas, como la protección de la libertad de expresión, a agentes privados que no rinden cuentas ante nadie. A pesar de haber adoptado medidas para hacer más transparentes sus normas y su relación con los Gobiernos, las empresas siguen siendo unos reguladores enigmáticos, que establecen una especie de “ley de las plataformas” en la que es difícil percibir elementos como claridad, coherencia, rendición de cuentas y reparación. Las Naciones Unidas, las organizaciones regionales y los órganos creados en virtud de los tratados de derechos humanos han afirmado que los derechos fuera de línea se aplican igualmente en línea, pero no siempre está claro que las empresas protejan los derechos de sus usuarios o que los Estados les ofrezcan en las leyes incentivos para hacerlo.

2. En el presente informe, el Relator Especial propone un marco para la moderación del contenido en línea generado por los usuarios que sitúe los derechos humanos en el centro mismo de la cuestión[2]. El Relator Especial pretende responder algunas preguntas básicas: ¿Qué responsabilidad tienen las empresas de asegurarse de que sus plataformas no interfieran con los derechos garantizados en virtud del derecho internacional? ¿Qué normas deben aplicarse en la moderación de los contenidos? ¿Deben los Estados regular la moderación del contenido comercial? y, en caso afirmativo, ¿De qué manera? La ley confía en que la transparencia y la responsabilidad de los Estados mitiguen las amenazas a la libertad de expresión. ¿Debemos esperar lo mismo de los agentes privados? ¿Qué forma adoptan los procesos de protección y reparación en la era digital?

3. En informes anteriores se han abordado algunas de esas cuestiones[3]. El presente informe se centra en la regulación del contenido generado por los usuarios, principalmente por parte de los Estados y las empresas que mantienen las redes sociales, pero de una manera que sea aplicable a todos los agentes pertinentes en el sector de la tecnología de la información y las comunicaciones (TIC). El Relator Especial esboza el marco jurídico de derechos humanos aplicable y describe los enfoques de la regulación de los contenidos que adoptan las empresas y los Estados. Propone normas y procesos que las empresas deberían adoptar para regular el contenido de manera compatible con las normas de derechos humanos.

4. Las bases iniciales para la elaboración del informe se asentaron en la investigación de las condiciones de servicio de las empresas, la presentación de informes de transparencia y otras fuentes secundarias. La iniciativa de solicitar comentarios dio lugar a 21 comunicaciones de los Estados y 29 de agentes no estatales (entre ellos 1 empresa). El Relator Especial visitó varias empresas de Silicon Valley y mantuvo conversaciones con otras en un esfuerzo por comprender sus enfoques de la moderación del contenido[4]. Hizo uso de las consultas con la sociedad civil celebradas en Bangkok y Ginebra en 2017 y 2018

[1] John Perry Barlow, “Una declaración de independencia del ciberespacio”, 8 de febrero de 1996.

[2] Por “moderación” se entiende el proceso mediante el cual las empresas de Internet determinan si el contenido generado por los usuarios se ajusta a las reglas establecidas en sus condiciones de servicio y otras normas.

[3] A/HRC/35/22 y A/HRC/32/38.

[4] El Relator Especial visitó las sedes de Facebook, Github, Google, Reddit y Twitter y mantuvo conversaciones con representantes de Yahoo/Oath, Line y Microsoft. También visitó la Fundación Wikimedia sin fines de lucro. Para seguir desarrollando las labores relacionadas con el presente informe, confía en poder visitar empresas ubicadas en Beijing, Moscú, Seúl y Tokio.

y de los debates en línea que mantuvo con expertos de América Latina, Oriente Medio y el Norte de África y África Subsahariana en 2018[5].

II. Marco jurídico

5. Las actividades de las empresas del sector de las TIC afectan, entre otros, a los derechos a la privacidad y a la participación pública y a las libertades de religión y de creencias, de opinión y de expresión, de reunión y de asociación. El presente informe se centra en la libertad de expresión, sin que por ello se deje de reconocer la interdependencia de los derechos, como la importancia de la privacidad como puerta de acceso a la libertad de expresión[6]. En el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos se establecen normas de ámbito mundial, ratificadas por 170 Estados, que se hacen eco de la Declaración Universal de Derechos Humanos al garantizar "el derecho de toda persona a no ser molestada a causa de sus opiniones" y "a buscar, recibir y difundir informaciones e ideas de toda índole, sin consideración de fronteras y por cualquier procedimiento"[7].

A. Obligaciones de los Estados

6. El derecho de los derechos humanos impone a los Estados la obligación de proporcionar un entorno propicio para la libertad de expresión y proteger el ejercicio de esa libertad. La obligación de garantizar la libertad de expresión obliga a los Estados a promover, entre otras cosas, la diversidad y la independencia de los medios de comunicación y el acceso a la información[8]. Además, los organismos internacionales y regionales han instado a los Estados a que promuevan el acceso universal a Internet[9]. Los Estados también tienen la obligación de velar por que las entidades privadas no interfieran con la libertad de opinión y de expresión[10]. En el tercero de los Principios Rectores sobre las Empresas y los Derechos Humanos, aprobados por el Consejo de Derechos Humanos en 2011, se hace hincapié en la obligación de los Estados de proporcionar un entorno que favorezca el respeto de los derechos humanos por parte de las empresas[11].

7. Los Estados no pueden restringir el derecho de las personas a no ser molestadas a causa de sus opiniones. En virtud de lo dispuesto en el artículo 19, párrafo 3, del Pacto, las limitaciones que se impongan a la libertad de expresión deben cumplir las siguientes condiciones bien establecidas:

- *Legalidad.* Las restricciones deben estar "previstas en la ley". En particular, deben ser aprobadas por los procedimientos jurídicos ordinarios y limitar la discrecionalidad del Gobierno de manera que se distinga entre las expresiones lícitas e ilícitas con "suficientemente precisión". Las restricciones impuestas en secreto no satisfacen esa exigencia fundamental[12]. La garantía de legalidad debe conllevar

[5] El Relator Especial desea expresar su agradecimiento a su asesor jurídico, Amos Toh, y a los alumnos de la International Justice Clinic de la Facultad de Derecho de la Universidad de California, en Irvine.

[6] Véase A/HRC/29/32, párrs. 16 a 18.

[7] Véanse la Carta Africana de Derechos Humanos y de los Pueblos, art. 9; la Convención Americana sobre Derechos Humanos, art. 13; y el Convenio para la Protección de los Derechos Humanos y las Libertades Fundamentales, art. 10. Véase también la comunicación del Centro de Estudios en Libertad de Expresión y Acceso a la Información.

[8] Declaración conjunta sobre la libertad de expresión y las "noticias falsas", la desinformación y la propaganda, 3 de marzo de 2017, secc. 3. Véase también la observación general núm. 34 (2011) del Comité de Derechos Humanos, sobre la libertad de opinión y de expresión, párrs. 18 y 40; A/HRC/29/32, párr. 61 y A/HRC/32/38, párr. 86.

[9] Véase la resolución 32/13 del Consejo de Derechos Humanos, párr. 12. Oficina del Relator Especial para la libertad de expresión de la Comisión Interamericana de Derechos Humanos, *Estándares para una internet libre, abierta e incluyente* (2016), párr 18.

[10] Véase la observación general núm. 34, párr. 7.

[11] A/HRC/17/31.

[12] *Ibid.*, párr. 25. A/HRC/29/32.

generalmente una supervisión por parte de las autoridades judiciales independientes[13].

- *Necesidad y proporcionalidad.* Los Estados deben demostrar que con la restricción se impone la menor carga posible al ejercicio del derecho y se protege, o es probable que se proteja, el interés legítimo del Estado en cuestión. Los Estados no pueden limitarse a alegar la necesidad a la hora de promulgar leyes restrictivas e imponer la restricción de una expresión concreta, sino que deben demostrarla[14].

- *Legitimidad.* Para que una restricción sea legal, debe estar orientada a proteger alguno de los intereses enumerados en el artículo 19, párrafo 3: los derechos o la reputación de los demás, la seguridad nacional o el orden público, o la salud o la moral públicas. Las restricciones destinadas a proteger los derechos de los demás, por ejemplo, comprenden "los derechos humanos reconocidos en el Pacto y, más en general, en la normativa internacional de los derechos humanos"[15]. Las restricciones destinadas a proteger los derechos a la privacidad, la vida, las debidas garantías procesales, la asociación y la participación en los asuntos públicos, por nombrar solo algunos, serían legítimas si se demostrase que cumplen los criterios de legalidad y necesidad. El Comité de Derechos Humanos advierte de que las restricciones tendientes a proteger la "moral pública" no deben derivarse "exclusivamente de una sola tradición", y debe procurarse que la restricción refleje la universalidad de los derechos humanos y el principio de no discriminación[16].

8. Las restricciones impuestas con arreglo al artículo 20, párrafo 2, del Pacto, en virtud del cual se exige a los Estados que prohíban "la apología del odio nacional, racial o religioso que constituya incitación a la discriminación, la hostilidad o la violencia", también deben satisfacer simultáneamente las condiciones de legalidad, necesidad y legitimidad[17].

B. Responsabilidad de las empresas

9. Las empresas de Internet se han convertido en plataformas fundamentales para la discusión y el debate, el acceso a la información, el comercio y el desarrollo humano[18]. Reúnen y conservan los datos personales de miles de millones de personas, incluso información sobre sus hábitos, sus movimientos y sus actividades, y a menudo afirman desempeñar funciones de carácter cívico. En 2004, Google manifestó su ambición de hacer "cosas buenas para el mundo incluso renunciando a algunos beneficios a corto plazo"[19]. El fundador de Facebook ha proclamado el deseo de "desarrollar la infraestructura social para dar a las personas el poder de construir una comunidad mundial que nos sirva a todos"[20]. Twitter ha prometido aplicar políticas que "mejoren —y no perjudiquen— una conversación libre y mundial"[21]. VKontakte, una empresa de comunicación social de la Federación de Rusia, dice que "une a personas de todo el mundo", en tanto que Tencent refleja el discurso del Gobierno de China cuando señala su objetivo de "contribuir a crear una sociedad armoniosa y convertirse en una empresa ejemplo de ciudadanía"[22].

[13] *Ibid.*
[14] Véase la observación general núm. 34, párr. 27.
[15] *Ibid.*, párr. 28.
[16] *Ibid.*, párr. 32.
[17] *Ibid.*, párr. 50. Véase también A/67/357.
[18] Véase, por ejemplo, Tribunal Supremo de los Estados Unidos, *Packingham v. North Carolina*, sentencia de 19 de junio de 2017; Tribunal Europeo de Derechos Humanos, *Times Newspapers Ltd. (Nos. 1 and 2) v. The United Kingdom* (demandas núms. 3002/03 y 23676/03), sentencia de 10 de marzo de 2009, párr 27.
[19] Declaración en el registro de valores (S-1) en virtud de la Ley de Valores de 1933, 18 de agosto de 2004.
[20] Mark Zuckerberg, "Building global community", Facebook, 16 de febrero de 2017.
[21] Twitter, declaración en el registro de valores S-1, 13 de octubre de 2013, págs. 91 y 92.
[22] VKontakte, información de la empresa; Tencent, "About Tencent".

10. Pocas empresas tienen en cuenta los principios de derechos humanos en sus operaciones, y la mayoría de las que lo hacen consideran limitada su capacidad de respuesta a las amenazas y las exigencias de los Gobiernos[23]. Sin embargo, en los Principios Rectores sobre las Empresas y los Derechos Humanos se establecen "normas de conducta mundial aplicables a todas las empresas" que deben regir todas las operaciones de las empresas cualquiera que sea el lugar en el que operen[24]. Si bien los Principios Rectores no son vinculantes, el importantísimo papel que desempeñan las empresas en la vida pública a nivel mundial aboga claramente en favor de su adopción y aplicación.

11. En los Principios Rectores se establece un marco en virtud del cual las empresas deben, como mínimo:

a) Abstenerse de causar o contribuir a cualquier consecuencia negativa sobre los derechos humanos y tratar de prevenir o mitigar esas consecuencias cuando estén directamente relacionadas con las operaciones, los productos o los servicios derivados de sus relaciones comerciales, incluso si no han contribuido a generarlas (principio 13);

b) Adoptar compromisos de política de alto nivel orientados a respetar los derechos humanos de sus usuarios (principio 16);

c) Llevar a cabo actividades de diligencia debida con las que se identifiquen, aborden y se dé cuenta de las posibles repercusiones de sus actividades en los derechos humanos, en particular mediante la realización periódica de evaluaciones de los riesgos y los efectos, consultas sustantivas con los grupos que pudieran resultar afectados y otras partes interesadas, así como medidas de seguimiento que sirvan para prevenir o mitigar esas repercusiones (principios 17 a 19);

d) Participar en estrategias de prevención y mitigación que respeten en la mayor medida posible los principios de derechos humanos internacionalmente reconocidos cuando se encuentren con requisitos incompatibles establecidos en la legislación local (principio 23);

e) Mantener en continuo examen sus esfuerzos por respetar los derechos, incluso mediante la celebración de consultas periódicas con los interesados y una comunicación frecuente, accesible y efectiva con los grupos afectados y el público en general (principios 20 y 21);

f) Proporcionar una reparación apropiada, incluso mediante mecanismos de solución de diferencias a nivel operacional a los que los usuarios puedan acceder sin acrecentar su "sensación de impotencia" (principios 22, 29 y 31).

III. Principales preocupaciones con respecto a la regulación del contenido

12. Los Gobiernos intentan controlar el entorno en el que las empresas moderan el contenido, mientras que las empresas defienden el acceso individual a sus plataformas mediante acuerdos de uso con unas condiciones de servicio en las que se determina qué se puede expresar y de qué forma puede expresarse.

A. Normativa de los Gobiernos

13. Los Estados exigen periódicamente a las empresas que restrinjan el contenido manifiestamente ilegal, como la representación de abusos sexuales de niños, las amenazas directas y verosímiles de causar un daño y la incitación a la violencia, dando por sentado que satisfacen con ello las condiciones de legalidad y necesidad[25]. Algunos Estados van

[23] Comunicación del Instituto de Derechos Humanos de Dinamarca. Cf. comunicación de Yahoo/Oath, 2016.

[24] Principios Rectores, principio 11.

[25] Irlanda ha establecido mecanismos conjuntos de regulación con las empresas para restringir las imágenes de abusos sexuales de menores: comunicación de Irlanda. Muchas empresas utilizan un

mucho más allá y recurren a la censura y a la aplicación de la ley para configurar el entorno reglamentario de la actividad en línea[26]. Unas leyes restrictivas con una redacción de carácter muy general sobre el "extremismo", la blasfemia, la difamación, el discurso "ofensivo" las "noticias falsas" y la "propaganda" suelen servir como pretexto para exigir que las empresas supriman la expresión legítima[27]. Cada vez con mayor frecuencia, los Estados apuntan específicamente al contenido de las plataformas en línea[28]. Otras leyes pueden interferir con la privacidad en línea de formas que disuadan a los usuarios de ejercer el derecho a la libertad de opinión y de expresión[29]. Muchos Estados también utilizan instrumentos de desinformación y propaganda para limitar la accesibilidad y la fiabilidad de los medios de comunicación independientes[30].

14. *Protección frente a la responsabilidad.* Desde el principio de la era digital, muchos Estados aprobaron normas para proteger a los intermediarios frente a la responsabilidad por el contenido que terceras partes pudieran publicar en sus plataformas. En la directiva de la Unión Europea sobre comercio electrónico, por ejemplo, se establece un régimen jurídico para proteger a los intermediarios frente a la responsabilidad por el contenido, salvo cuando vayan más allá de su papel como simple "canal", "depósito" o "anfitrión" de la información facilitada por los usuarios[31]. En el artículo 230 de la Ley de Decencia de las Comunicaciones de los Estados Unidos se prevé, en términos generales, la inmunidad de los proveedores de "servicios informáticos interactivos" que acogen o publican información acerca de los demás, aunque desde entonces esa inmunidad se ha reducido[32]. En virtud del régimen de responsabilidad de los intermediarios en el Brasil se requiere una orden judicial para restringir el contenido particular[33], mientras que en el régimen de responsabilidad de los intermediarios de la India se establece un procedimiento de "notificación y retirada" que supone la necesidad de que medie la orden de un juzgado u otro órgano judicial[34]. En los Principios de Manila de 2014 sobre la responsabilidad de los intermediarios, elaborados por una coalición de expertos de la sociedad civil, se establecen los principios esenciales que deben guiar cualquier marco rector de la responsabilidad de los intermediarios.

15. *Imposición de obligaciones a las empresas.* Algunos Estados imponen a las empresas la obligación de restringir el contenido con arreglo a criterios jurídicos vagos o complejos, sin un examen judicial previo y con la amenaza de sanciones severas. Por ejemplo, en la Ley de Ciberseguridad de China, de 2016, se imponen vagas prohibiciones contra la difusión de información "falsa" que perturbe el "orden social o económico", o la unidad o la seguridad nacionales. También se obliga a las empresas a supervisar sus redes y denunciar a las autoridades cualquier infracción[35]. Al parecer, el incumplimiento de esa Ley

algoritmo de reconocimiento de imágenes para detectar y eliminar las imágenes de pornografía infantil: comunicaciones del Open Technology Institute, pág. 2 y ARTICLE 19, pág. 8.

[26] Véase A/HRC/32/38, párrs. 46 y 47. Sobre los apagones de Internet, véase A/HRC/35/22, párrs. 8 a 16 y los ejemplos de las comunicaciones del Relator Especial: núms. UA TGO 1/2017, UA IND 7/2017 y AL GMB 1/2017.

[27] Comunicaciones núms. OC MYS 1/2018; UA RUS 7/2017; UA ARE 7/2017, AL-BHR 8/2016, AL SGP 5/2016 y OL RUS 7/2016. Azerbaiyán prohíbe la propaganda del terrorismo, el extremismo religioso y el suicidio: comunicación de Azerbaiyán.

[28] Véanse las comunicaciones núms. OL PAK 8/2016 y OL LAO 1/2014; Asociación para el Progreso de las Comunicaciones, *Unshackling Expression: A Study on Laws Criminalising Expression Online in Asia*, GISWatch 2017 Special Edition.

[29] A/HRC/29/32.

[30] Véase, por ejemplo, Gary King, Jennifer Pan y Margaret E. Roberts, "How the Chinese Government fabricates social media posts for strategic distraction, not engaged argument", *American Political Science Review*, vol. 111, núm. 3 (2017), págs. 484 a 501.

[31] Directiva núm. 2000/31/CE del Parlamento Europeo y del Consejo, de 8 de junio de 2000.

[32] Código 47 de los Estados Unidos, párr. 230. Véase también la Ley para Permitir a los Estados y las Víctimas Combatir el Tráfico Sexual en Línea (H.R. 1865).

[33] *Marco Civil da Internet*, Ley Federal 12965, arts. 18 y 19.

[34] Tribunal Supremo de la India, *Shreya Singhal v. Union of India*, sentencia de 24 de marzo de 2015.

[35] Artículos 12 y 47. Comunicación de Human Rights in China, 2016, pág. 12. Pueden consultarse observaciones sobre una versión anterior de la Ley de Ciberseguridad en la comunicación núm. OLC CHN 7/2015. Véase también Global Voices "Netizen Report: Internet censorship bill looms large over Egypt", 16 de marzo de 2018; República de Sudáfrica, proyecto de enmienda de la Ley de Cinematografía y Publicaciones (B 61-2003).

ha dado lugar a la imposición de cuantiosas multas a las principales plataformas de medios sociales[36].

16. La obligación de vigilar y eliminar rápidamente el contenido inapropiado generado por los usuarios también ha aumentado en todo el mundo, con el establecimiento de marcos sancionadores que pueden socavar la libertad de expresión, incluso en sociedades democráticas. En virtud de la Ley de Vigilancia de la Red *(NetzDG)* de Alemania se exige a las grandes empresas de medios sociales que eliminen el contenido incompatible con determinadas leyes locales en plazos muy breves, con importantes sanciones por incumplimiento[37]. La Comisión Europea incluso ha recomendado a sus Estados miembros que establezcan obligaciones jurídicas para la vigilancia activa y el filtrado de los contenidos ilegales[38]. En las directrices sobre la difusión de contenido por los medios sociales durante las elecciones, aprobadas en Kenya en 2017, se exige a las plataformas que "cierren las cuentas utilizadas para difundir contenido político indeseable en sus plataformas" en un plazo de 24 horas[39].

17. Teniendo en cuenta las preocupaciones legítimas del Estado, como el derecho a la privacidad y la seguridad nacional, el atractivo que despierta la regulación es comprensible. Sin embargo, esas normas entrañan riesgos para la libertad de expresión y ejercen una presión considerable sobre las empresas que puede llevarlas a eliminar incluso contenidos lícitos en un afán desmedido por evitar la responsabilidad. También suponen la delegación de funciones normativas a agentes privados si la existencia de unos instrumentos básicos de rendición de cuentas. La exigencia de la eliminación de contenido de forma rápida y automática lleva consigo el riesgo de que aparezcan nuevas formas de censura previa que ya amenazan los esfuerzos creativos en el contexto de los derechos de autor[40]. Las cuestiones complejas de hecho y de derecho deberían ser resueltas por las instituciones públicas, no por agentes privados cuyos procedimientos actuales tal vez no sean compatibles con las normas relativas a las debidas garantías procesales y cuya motivación es principalmente económica[41].

18. *Supresión de contenido a nivel mundial.* Algunos Estados exigen la eliminación extraterritorial de vínculos, sitios web y otros contenidos que supuestamente vulneren la legislación local[42]. Esas exigencias plantean la grave preocupación de que los Estados puedan interferir con el derecho a la libertad de expresión "sin consideración de fronteras". La lógica de esas exigencias permitiría ejercer la censura a través de las fronteras, en beneficio de quienes la aplican de forma más restrictiva. Debe exigirse a quienes soliciten ese tipo de eliminación de contenidos que formulen su solicitud en todas las jurisdicciones que corresponda, recurriendo para ello al procedimiento judicial ordinario.

[36] PEN América, *Forbidden Feeds: Government Controls on Social Media in China* (2018), pág. 21.

[37] Ley para Mejorar la Aplicación de la Ley en las Redes Sociales (Ley de Vigilancia de la Red), julio de 2017. Véase la comunicación núm. OL DEU 1/2017.

[38] Comisión Europea, recomendación sobre medidas para abordar de manera eficaz el contenido ilegal en línea (última actualización: 5 de marzo de 2018).

[39] Véase la comunicación núm. OL KEN 10/2017; Javier Pallero, "Libertad de expresión en peligro en Honduras", Access Now, 12 de febrero de 2018.

[40] Véase Comisión Europea, propuesta de directiva del Parlamento Europeo y del Consejo sobre los derechos de autor en el mercado único digital, COM (2016) 593 final, art. 13; Daphne Keller, "Problems with filters in the European Commission's platforms proposal", Stanford Law School Center for Internet and Society, 5 de octubre de 2017; Comunicación de la Fundación Karisma, 2016, págs. 4 a 6.

[41] Según la legislación de la Unión Europea, los motores de búsqueda deben determinar la validez de las solicitudes formuladas al amparo del "derecho al olvido". Tribunal de Justicia de la Unión Europea, *Google Spain c. la Agencia Española de Protección de Datos y Mario Costeja González* (asunto C-131/12), sentencia (Gran Sala) de 13 de mayo de 2014; Comunicaciones de ARTICLE 19, págs. 2 y 3 y Access Now, págs. 6 y 7; Google, "Updating our 'right to be forgotten' Transparency Report"; Theo. Bertram y otros, *Three Years of the Right to be Forgotten* (Google, 2018).

[42] Véanse, por ejemplo, PEN, *Forbidden Feeds*, págs. 36 y 37; Tribunal Supremo del Canadá, *Google Inc c. Equuestek Solutions Inc.*, sentencia de 28 de junio de 2017; Tribunal de Justicia de la Unión Europea, *Google Inc c. Commission nationale de l'Informatique et des Libertés (CNIL)* (asunto C-507/17); comunicación de Global Network Initiative, pág. 6.

19. *Exigencias de los Gobiernos que no tienen sustento en la legislación nacional.* Las empresas distinguen entre las solicitudes de retirada de contenido presuntamente ilegal presentadas por la vía ordinaria y las solicitudes de eliminación basadas en sus propias condiciones de servicio[43]. (Las eliminaciones de carácter legal por lo general se aplican únicamente en la jurisdicción requirente, mientras que las eliminaciones basadas en las condiciones de servicio se suelen aplicar en todo el mundo.) Las autoridades estatales solicitan cada vez con mayor frecuencia la eliminación de contenidos al margen del proceso judicial, o incluso a través de peticiones basadas en las condiciones de servicio[44]. Varias autoridades han establecido dependencias especializadas en señalar a las empresas el contenido que deben eliminar. La Unidad de Notificación de Contenidos de Internet de la Unión Europea, por ejemplo, "alerta del contenido en línea relacionado con el terrorismo y el extremismo violento y coopera con los proveedores de servicios en línea con el fin de eliminar ese contenido"[45]. Australia cuenta con mecanismos similares[46]. En Asia Sudoriental, se ha informado de que partidos aliados con los Gobiernos al parecer intentan utilizar las solicitudes basadas en las condiciones de servicio para restringir las críticas de carácter político[47].

20. Los Estados también ejercen presión sobre las empresas para que aceleren la eliminación de contenidos mediante normas no vinculantes, la mayoría de las cuales son de dudosa transparencia. Tres años de prohibición de YouTube en el Pakistán animaron a Google a crear una versión local más receptiva a las exigencias del Gobierno relativas a la eliminación de contenido "ofensivo"[48]. Se dijo que Facebook e Israel acordaron coordinar sus esfuerzos y su personal para vigilar y eliminar la "incitación" en línea. Los detalles de ese acuerdo no se divulgaron, pero el Ministro de Justicia de Israel afirmó que entre junio y septiembre de 2016, Facebook aceptó casi todas las solicitudes del Gobierno para la eliminación de contenidos relacionados con la "incitación"[49]. Los arreglos para coordinar las medidas contra los contenidos provenientes de los Estados aumentan la preocupación por la posibilidad de que las empresas desempeñen funciones públicas sin la supervisión de los tribunales y otros mecanismos de rendición de cuentas[50].

21. El Código de Conducta de la Unión Europea de 2016 sobre la lucha contra el discurso ilegal de odio en línea es fruto de un acuerdo para eliminar contenidos concertado entre la Unión Europea y cuatro grandes empresas, que se comprometen a colaborar con "detectores fiables" y promover "argumentos contrarios independientes"[51]. Si bien el fomento de los argumentos contrarios puede resultar atractivo ante el contenido "extremista" o "terrorista", la presión para adoptar esos enfoques conlleva el riesgo de transformar las plataformas en transmisores de propaganda más allá de los ámbitos de interés legítimo[52].

[43] Compárese Twitter Transparency Report: Removal Requests (enero a junio de 2017) con Twitter Transparency Report: Government Terms of Service Reports (enero a junio de 2017). Véase también Facebook, Government requests: Frequently Asked Questions (FAQs).

[44] Comunicaciones de ARTICLE 19, pág. 2 y Global Network Initiative, pág. 5.

[45] Unidad de Notificación de Contenidos de Internet de la Unión Europea, informe sobre el primer año, secc. 4.11; comunicaciones de European Digital Rights (EDRi), pág. 1 y Access Now, págs. 2 y 3.

[46] Comunicación de Australia.

[47] Southeast Asian Press Alliance, pág. 1.

[48] Comunicación de Digital Rights Foundation.

[49] Comunicación de 7amleh - Arab Center for the Advancement of Social Media.

[50] Asociación para el Progreso de las Comunicaciones, pág. 14 y 7amleh.

[51] "Detectores fiables es una condición que se otorga a ciertas organizaciones que les permite denunciar contenidos ilícitos a través de un sistema o canal de información especial, que no está a disposición de los usuarios corrientes." Comisión Europea, Código de Conducta sobre la lucha contra el discurso ilegal de odio en línea: primeros resultados sobre su aplicación (diciembre de 2016).

[52] Las mismas empresas crearon el Foro Mundial de Internet para Contrarrestar el Terrorismo, una iniciativa cuyo objeto es elaborar instrumentos tecnológicos a nivel de todo el sector para eliminar el contenido terrorista en sus plataformas. Google, "Update on the Global Internet Forum to Counter Terrorism", 4 de diciembre de 2017.

B. Moderación del contenido por las empresas

Cumplimiento de la legislación nacional por las empresas

22. Toda empresa se compromete, en principio, a cumplir la ley en el curso de su actividad. En palabras de Facebook: "Si, tras un cuidadoso examen jurídico, determinamos que el contenido es ilegal con arreglo la legislación local, lo retiramos en el país o territorio de que se trate"[53]. Tencent, el propietario de la aplicación de charla por el móvil y red social WeChat, va considerablemente más allá al exigir a cualquiera que utilice la plataforma en China y a los ciudadanos chinos que la utilicen "en cualquier parte del mundo" que cumplan con unas restricciones de contenido que replican la legislación o la política de China[54]. Varias empresas también colaboran entre sí y con los órganos normativos para eliminar las imágenes de abusos sexuales de niños[55].

23. El compromiso de cumplir la ley puede ser complicado cuando la legislación del Estado de que se trate es vaga, está sujeta a diversas interpretaciones o es incompatible con las normas de derechos humanos. Por ejemplo, las leyes contra el "extremismo" en las que no se define el término clave dan a las autoridades gubernamentales discrecionalidad a la hora de presionar a las empresas para que eliminen contenidos por motivos cuestionables[56]. De igual manera, las empresas suelen ser objeto de presiones para que cumplan leyes estatales en virtud de las cuales se penaliza el contenido que se considera, por ejemplo, falso, blasfemo, crítico con el Estado o difamatorio para los funcionarios públicos. Como se explica más adelante, los Principios Rectores proporcionan instrumentos para minimizar el impacto de esas leyes en los usuarios particulares. La Global Network Initiative, una iniciativa de múltiples interesados que ayuda a las empresas del sector de las TIC a sortear los desafíos que plantean los derechos humanos, ha elaborado nuevas directrices sobre la forma de utilizar esos instrumentos[57]. Un medio para minimizar los problemas es la transparencia: muchas empresas informan anualmente del número de solicitudes que reciben y ejecutan en cada Estado[58]. Sin embargo, las empresas no siempre revelan información suficiente sobre cómo responden a las solicitudes de los Gobiernos, ni tampoco informan periódicamente sobre las solicitudes de los Gobiernos formuladas al amparo de las condiciones de servicio[59].

[53] Facebook, Government requests: FAQs. Véanse también las solicitudes legales de retirada de Google; Normas y políticas de Twitter; política de Reddit sobre el contenido.

[54] Tencent, condiciones de servicio: introducción; Tencent, acuerdo sobre licencias de programas informáticos y servicios de Tencent Wenxin.

[55] Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, *Fostering Freedom Online: The Role of Internet Intermediaries* (París, 2014), págs. 56 y 57.

[56] Véase Maria Kravchenko, "Inappropriate enforcement of anti-extremist legislation in Russia in 2016", Centro de Información y Análisis SOVA, 21 de abril de 2017; Danielle Citron, "Extremist speech, compelled conformity, and censorship creep", *Notre Dame Law Review*, vol. 93, núm. 3 (2018), págs. 1035 a 1071.

[57] Global Network Initiative, Principles on Freedom of Expression and Privacy, secc. 2. Entre las empresas relacionadas con los medios sociales que participan en la Iniciativa se encuentran Facebook, Google, Microsoft/LinkedIn y Yahoo/Oath.

[58] Véase el párrafo 39. Además, Automattic, Google, Microsoft/Bing y Twitter se encuentran entre las empresas que periódicamente, aunque no necesariamente de forma exhaustiva, publican en la base de datos Lumen las solicitudes de los Gobiernos relativas a la propiedad intelectual y a la retirada de contenidos.

[59] Ranking Digital Rights, 2017 Índice de Responsabilidad Empresarial, pág. 28.

Normas de moderación de las empresas

24. Las empresas de Internet exigen a sus usuarios que respeten las condiciones de servicio y "las normas comunitarias" que rigen la libertad de expresión en sus plataformas[60]. En las condiciones de servicio de las empresas, que los usuarios están obligados a aceptar para poder utilizar la plataforma, se establecen cuáles son las jurisdicciones para la solución de controversias y se reservan la discrecionalidad en cuanto a la adopción de medidas sobre el contenido y sobre la cuenta[61]. Las políticas sobre el contenido de la plataforma son un subconjunto de esas condiciones, en las que se exponen las restricciones acerca de lo que los usuarios pueden expresar y cómo pueden hacerlo. La mayoría de las empresas no basan expresamente las normas sobre el contenido en un ordenamiento jurídico concreto que pueda regular la libertad de expresión, como la legislación nacional o el derecho internacional de los derechos humanos. Sin embargo, el gigante chino de la búsqueda Baidu prohíbe el contenido que "se oponga a los principios básicos establecidos en la Constitución" de la República Popular China[62].

25. La elaboración de políticas de moderación de contenido suele entrañar la participación de un abogado, directores de producto y políticas públicas y altos ejecutivos. Las empresas pueden establecer equipos de "confianza y seguridad" para hacer frente a los mensajes basura, el fraude y los abusos, y equipos de lucha contra el terrorismo que pueden ocuparse de abordar el contenido de naturaleza terrorista[63]. Algunas empresas han elaborado mecanismos para recabar la participación de grupos externos sobre aspectos especializados de las políticas sobre el contenido[64]. El aumento exponencial del contenido generado por los usuarios ha dado lugar a la elaboración de normas detalladas y en constante evolución. Esas normas varían en función de una serie de factores, desde el tamaño de la empresa, los ingresos y el modelo de negocio hasta la "marca y la reputación de la plataforma, su tolerancia al riesgo y el tipo de participación de los usuarios que desea atraer"[65].

Esferas de preocupación en el ámbito de las normas sobre el contenido

26. *Vaguedad de las normas.* Las prohibiciones que se imponen a las empresas de ensalzar o promover el terrorismo[66], apoyar o elogiar a los dirigentes de organizaciones peligrosas[67] y albergar contenidos que promuevan actos de terrorismo o inciten a la violencia[68], son al igual que la legislación contra el terrorismo, excesivamente vagas[69]. En las políticas de las empresas con respecto al odio, el acoso y los abusos no se indica claramente qué constituye un delito. La prohibición de Twitter de mostrar una "conducta que incite al miedo a un grupo protegido" y la distinción que hace Facebook entre "ataques directos" contra características protegidas y "el contenido simplemente desagradable u ofensivo" son fundamentos subjetivos e inestables en que basar la moderación del contenido[70].

[60] Jamila Venturini y otros, *Terms of Service and Human Rights: An Analysis of Online Platform Contracts* (Río de Janeiro, Revan, 2016).

[61] Acuerdo de usuario de Baidu ("Cualquier contenido podrá ser eliminado y suprimido por cualquier motivo a discreción de Baidu"); condiciones de servicio de Tencent ("Nos reservamos el derecho a bloquear o eliminar su contenido por cualquier motivo, incluso si, a nuestro juicio, es apropiado hacerlo o si así lo requieren las leyes y reglamentos aplicables"); condiciones de servicio de Twitter ("Podemos suspender o clausurar su cuenta o dejar de proporcionarle la totalidad o una parte de los servicios en cualquier momento y por cualquier motivo o incluso sin motivo").

[62] Condiciones de servicio de Baidu, secc. 3.1.

[63] Monika Bickert, "Hard questions: how we counter terrorism", 15 de junio de 2017.

[64] Véanse, por ejemplo, Twitter Trust and Safety Council y YouTube Trusted Flagger Program.

[65] Sarah Roberts, *Content Moderation* (Universidad de California, Los Angeles, 2017). Véase también la comunicación de ARTICLE 19, pág. 2.

[66] Normas y políticas de Twitter (grupos extremistas violentos).

[67] Normas comunitarias de Facebook (organizaciones peligrosas).

[68] Políticas de YouTube (políticas sobre el contenido violento o explícito).

[69] Véase A/HRC/31/65, párr. 39.

[70] Normas comunitarias de Facebook (discurso de odio); Normas y políticas de Twitter (política sobre la conducta de odio).

27. *Odio, acoso, abuso.* La vaguedad de las políticas relativas al acoso y el discurso de odio ha dado lugar a denuncias de una aplicación incoherente de esas políticas que perjudica a las minorías, al tiempo que refuerza la situación de los grupos dominantes o poderosos. Los usuarios y la sociedad civil informan de actos de violencia y abuso contra la mujer, incluidas las amenazas físicas, los comentarios misóginos, la publicación de imágenes íntimas falsas o sin consentimiento y la publicación de información personal confidencial[71]; las amenazas de agresión contra los grupos políticamente marginados[72], las razas y las castas minoritarias[73] y los grupos étnicos que sufren persecución violenta[74]; y los abusos dirigidos contra los refugiados, los migrantes y los solicitantes de asilo[75]. Al mismo tiempo, las plataformas habrían reprimido el activismo en favor de las personas lesbianas, gais, bisexuales, transgénero y asexuadas[76]; la contestación contra los Gobiernos represivos[77]; la denuncia de la depuración étnica[78]; y las críticas de los fenómenos y las estructuras de poder de naturaleza racista[79].

28. La magnitud y la complejidad de la lucha contra las expresiones de odio plantea problemas a largo plazo y pueden llevar a las empresas a restringir esa expresión aun cuando no esté claramente vinculada a resultados adversos (ya que las actividades de fomento del odio están relacionadas con la incitación en el artículo 20 del Pacto Internacional de Derechos Civiles y Políticos). No obstante, las empresas deben explicar en qué se basan esas restricciones y demostrar la necesidad y la proporcionalidad de esas medidas (como la eliminación del contenido o la suspensión de la cuenta). Un nivel de transparencia significativo y coherente sobre la ejecución de las políticas relativas al discurso de odio mediante la información razonada sobre casos concretos puede proporcionar también un grado de conocimiento del fenómeno que ni siquiera las explicaciones más detalladas pueden ofrecer[80].

29. *Contexto.* Las empresas hacen hincapié en la importancia del contexto a la hora de valorar la aplicación de las restricciones de carácter general[81]. No obstante, la atención al contexto no ha impedido la retirada de imágenes de desnudos con valor histórico, cultural o educativo[82]; relatos históricos y documentales de conflictos[83]; pruebas de crímenes de guerra[84]; intervenciones en contra de los grupos que promueven el odio[85]; o esfuerzos por impugnar o denunciar el lenguaje racista, homófobo o xenófobo[86]. El examen razonado del contexto puede verse frustrado por las limitaciones de tiempo y de recursos de que adolecen los moderadores humanos, la dependencia excesiva de la automatización o la insuficiente comprensión de los matices lingüísticos y culturales[87]. Las empresas han instado a los

[71] Amnistía Internacional, *Toxic Twitter: A Toxic Place for Women*; Comunicación de la Asociación para el Progreso de las Comunicaciones, pág. 2.

[72] Comunicaciones de 7amleh y la Asociación para el Progreso de las Comunicaciones, pág. 15.

[73] Ijeoma Oluo, "Facebook's complicity in the silencing of black women", Medium, 2 de agosto de 2017; comunicaciones del Center for Communications Governance, pág. 5 y la Asociación para el Progreso de las Comunicaciones, págs. 11 y 12.

[74] Declaración del Relator Especial sobre la situación de los derechos humanos en Myanmar, Yanghee Lee, en el 37º período de sesiones del Consejo de Derechos Humanos, 12 de marzo de 2018.

[75] Comunicación de la Asociación para el Progreso de las Comunicaciones, pág. 12.

[76] Fundación de la Frontera Electrónica, pág. 5.

[77] *Ibid.*; comunicaciones de la Asociación para el Progreso de las Comunicaciones y 7amleh.

[78] Betsy Woodruff, "Facebook silences Rohingya reports of ethnic cleansing", *The Daily Beast*, 18 de septiembre de 2017; Comunicación de ARTICLE 19, pág. 9.

[79] Julia Angwin y Hannes Grasseger, "Facebook's secret censorship rules protect white men from hate speech but not black children", *ProPublica*, 28 de junio de 2017.

[80] Véanse los párrafos 52 y 62.

[81] Twitter, "Our approach to policy development and enforcement philosophy"; políticas de YouTube (la importancia del contexto); Richard Allan, "Hard questions: who should decide what is hate speech in an online global community?", Facebook Newsroom, 27 de junio de 2017.

[82] Comunicaciones de OBSERVACOM, pág. 11 y ARTICLE 19, pág. 6.

[83] Comunicación de WITNESS, págs. 6 y 7.

[84] *Ibid.*

[85] Comunicación de la Fundación de la Frontera Electrónica, pág. 5.

[86] Comunicación de la Asociación para el Progreso de las Comunicaciones, pág. 14.

[87] Véase Allan, "Hard questions".

usuarios a que complementen el contenido controvertido con detalles contextuales, pero la viabilidad y la eficacia de esas directrices no están claras[88].

30. *Requisito de proporcionar la identidad verdadera.* A fin de hacer frente a los abusos en línea, algunas empresas han impuesto requisitos de "identidad verdadera"[89]; otras abordan las cuestiones relativas a la identidad de manera más flexible[90]. La eficacia del requisito de proporcionar el nombre real como salvaguardia contra el abuso en línea es cuestionable[91]. De hecho, la insistencia estricta en que se proporcione el nombre real ha supuesto revelar la identidad de blogueros y activistas que utilizan seudónimos para protegerse, exponiéndolos así a graves peligros físicos[92]. También ha supuesto el bloqueo de las cuentas de usuarios y activistas en favor de los derechos de las personas lesbianas, gais, bisexuales, transexuales y asexuales, artistas travestidos y de las cuentas de usuarios con nombres que no son ingleses o que son poco convencionales[93]. Habida cuenta de que el anonimato en línea a menudo es necesario para la seguridad física de los usuarios vulnerables, los principios de derechos humanos se inclinan hacia la protección del anonimato, con sujeción únicamente a las limitaciones que conduzcan a proteger la identidad[94]. Unas normas estrictas en relación con la suplantación de identidad que limiten la capacidad de los usuarios para hacerse pasar por otra persona de forma maliciosa o engañosa puede ser un medio más apropiado para proteger la identidad, los derechos y la reputación de otros usuarios[95].

31. *Desinformación.* La desinformación y la propaganda dificultan el acceso a la información y merman la confianza del público en los medios de comunicación y las instituciones de Gobierno. Las empresas se enfrentan a una presión cada vez mayor para hacer frente a la propagación de la desinformación mediante enlaces a artículos o sitios web que muestran noticias ficticias de terceros, cuentas falsas, anuncios engañosos o la manipulación del orden de aparición en las búsquedas[96]. Sin embargo, como las medidas contundentes, como la eliminación o el bloqueo del sitio web, pueden dar lugar a una injerencia grave en la libertad de expresión, las empresas deben actuar cuidadosamente al elaborar las políticas relativas a la desinformación[97]. Las empresas han adoptado diversas respuestas, incluidos los acuerdos con terceros que se encargan de la verificación, la intensificación de la vigilancia del cumplimiento de las políticas de publicidad, una mayor vigilancia de las cuentas sospechosas, la introducción de cambios en la edición de contenidos y los algoritmos que dan lugar a la clasificación de las búsquedas, y la formación de los usuarios sobre la detección de información falsa[98]. Algunas medidas, en particular las que se centran en la restricción del contenido de noticias, pueden poner en peligro las fuentes de noticias independientes y alternativas o los contenidos satíricos[99]. Las autoridades gubernamentales han adoptado posiciones que pueden reflejar unas

[88] Políticas de YouTube (la importancia del contexto); Normas comunitarias de Facebook (discurso de odio).

[89] Normas comunitarias de Facebook (utilizar la verdadera identidad). Obsérvese que Facebook acepta ahora excepciones a su política sobre el nombre real que analiza caso por caso, pero incluso esa medida ha sido criticada por insuficiente: comunicación de Access Now, pág. 12. Baidu requiere incluso información que identifique al usuario: acuerdo de usuario de Baidu.

[90] Centro de ayuda de Twitter: "Ayuda con el registro del nombre de usuario"; Instagram, "Empezar a utilizar Instagram".

[91] J. Nathan Matias, "The real name fallacy", Coral Project, 3 de enero de 2017.

[92] Comunicación de Access Now, pág. 11.

[93] Dia Kayyali, "Facebook's name policy strikes again, this time at Native Americans", Fundación de la Frontera Electrónica, 13 de febrero de 2015.

[94] Véase A/HRC/29/32, párr. 9.

[95] Normas y políticas de Twitter (política sobre la suplantación).

[96] *Ibid.*; Allen Babajanian y Christine Wendel, "#FakeNews: innocuous or intolerable?", Wilton Park report 1542, abril de 2017.

[97] Declaración conjunta de 2017.

[98] Comunicaciones de la Asociación para el Progreso de las Comunicaciones, págs. 4 a 6 y ARTICLE 19, pág. 4.

[99] Comunicación de la Asociación para el Progreso de las Comunicaciones, pág. 5.

expectativas excesivamente optimistas acerca de la capacidad de la tecnología para resolver esos problemas por sí sola[100].

Procesos e instrumentos de las empresas para la moderación

32. *Señalización y eliminación automáticas, y filtrado previo a la publicación.* El enorme volumen de contenido generado por los usuarios ha llevado a las principales empresas a desarrollar instrumentos de moderación automatizados. La automatización se ha utilizado principalmente para señalar contenidos a la atención de revisores humanos y, a veces, para eliminarlo. La utilización de instrumentos automatizados para evitar la infracción de los derechos de autor en música e imágenes en el momento de su carga han planteado preocupaciones por el exceso de bloqueos, y las solicitudes de que el filtrado en el momento de la carga se aplique también a los contenidos relacionados con el terrorismo y otros tipos de contenido amenazan con el establecimiento de regímenes totalitarios y desproporcionados de censura previa a la publicación[101].

33. La automatización puede aportar valor a las empresas que tienen que valorar enormes volúmenes de contenido generado por los usuarios, y para ello se utilizan instrumentos que van desde los filtros de palabras y la detección de mensajes basura, hasta los algoritmos de comparación criptográfica y el procesamiento del lenguaje natural[102]. La comparación criptográfica se utiliza profusamente para detectar imágenes de abusos sexuales a niños, pero su aplicación al contenido "extremista" —que por lo general requiere la evaluación del contexto— es difícil sin la existencia de normas claras sobre el "extremismo" o la intervención de personas[103]. Lo mismo ocurre con el procesamiento del lenguaje natural[104].

34. *La detección por los usuarios y los detectores fiables.* La detección por los usuarios ofrece a estos la posibilidad de señalar contenidos inapropiados a la atención de los moderadores. Las denuncias normalmente no permiten un examen de los matices en cuanto a los límites lo que es apropiado (por ejemplo, por qué el contenido puede ser ofensivo, pero en el contexto general, es mejor dejarlo)[105]. También se han "aprovechado" para aumentar la presión sobre las plataformas para que eliminen contenidos en apoyo de las minorías sexuales y los musulmanes[106]. Muchas empresas han elaborado listas especializadas de detectores "de confianza", normalmente expertos, usuarios con gran influencia y a veces, al parecer, elementos vinculados al gobierno[107]. Hay poca o ninguna información publicada en la que se expliquen los procedimientos empleados para la selección de detectores especializados, su interpretación de las normas jurídicas o comunitarias o su influencia en las decisiones de las empresas.

35. *Evaluación por personas.* A menudo, la automatización se complementa con un examen realizado por personas, y las mayores empresas de redes sociales reúnen grandes equipos de moderadores para examinar el contenido señalado[108]. Ese contenido puede enviarse a los moderadores, que normalmente están facultados para tomar una decisión —a

[100] Véase la comunicación núm. OL ITA 1/2018; Cf. Comisión Europea, *A Multi-Dimensional Approach to Disinformation: Final Report of the Independent High-level Group on Fake News and Disinformation* (Luxemburgo, 2018).

[101] Al parecer, el Reino Unido de Gran Bretaña e Irlanda del Norte elaboró un instrumento para detectar y eliminar automáticamente el contenido terrorista en el momento de la carga. Ministerio del Interior, "New technology revealed to help fight terrorist content online", 13 de febrero de 2018.

[102] Center for Democracy and Technology, *Mixed Messages? The Limits of Automated Media Content Analysis* (noviembre de 2017), pág. 9.

[103] Comunicación del Open Technology Institute, pág. 2.

[104] Center for Democracy and Technology, *Mixed Messages?*,pág. 4.

[105] Sobre los avisos de los usuarios, véase en general Kate Crawford y Tarleton Gillespie, "What is a flag for? Social media reporting tools and the vocabulary of complaint", *New Media and Society*, vol. 18, núm. 3 (marzo de 2016), págs. 410 a 428.

[106] *Ibid.*, pág. 421.

[107] Ayuda de YouTube, Programa de detectores fiables de YouTube; Ayuda de YouTube, "Colabora con quienes contribuyen a YouTube".

[108] Véase Sarah Roberts, "Commercial content moderation: digital laborers' dirty work", *Media Studies Publications*, paper 12 (2016).

menudo en minutos— acerca de la idoneidad del contenido y eliminarlo o permitirlo. En los casos en que la pertinencia de un contenido concreto resulte difícil de determinar, los moderadores pueden remitir su examen a los equipos de revisión del contenido de la dirección de la empresa. A su vez, los empleados de la empresa —generalmente equipos relacionados con las políticas con el público o de "confianza y seguridad", con la participación del asesor general— serán los encargados de adoptar decisiones sobre la supresión. La información que las empresas divulgan acerca de los debates sobre la eliminación de contenido, en general o en casos concretos, es limitada[109].

36. *Adopción de medidas en relación con la cuenta o el contenido.* La existencia de contenidos inapropiados puede desencadenar una serie de actuaciones por parte de la empresa. Las empresas pueden limitar la eliminación de contenidos a una jurisdicción, a una serie de jurisdicciones, o a toda una plataforma o conjunto de plataformas. Pueden aplicar limitaciones de edad, advertencias o desmonetización[110]. Las infracciones pueden dar lugar a la suspensión temporal de cuenta, en tanto que la reincidencia puede dar lugar a su desactivación definitiva. En algunos casos, muy pocos, aparte del respeto de los derechos de autor las empresas cuentan con procedimientos de "notificación en contrario" que permiten a los usuarios la publicación de observaciones para impugnar la eliminación de contenido.

37. *Notificación.* Una queja habitual es que los usuarios que publican contenido denunciado, o las personas que denuncian abusos, a veces no reciben ninguna notificación de la decisión de eliminarlo u otra medida que se haya podido adoptar[111]. Incluso cuando las empresas envían notificaciones, lo habitual es que en ellas solamente se indiquen las medidas adoptadas y un motivo genérico para su adopción. Al menos una empresa ha intentado facilitar más información en sus notificaciones, pero no está claro si los detalles adicionales que se aportan constituyen una explicación suficiente en todos los casos[112]. La transparencia y las notificaciones van de la mano: una trasparencia consolidada a nivel operacional que facilite el conocimiento del usuario acerca del enfoque de la plataforma con respecto a la supresión del contenido alivia la presión sobre la información facilitada en las notificaciones en los casos individuales, mientras que el debilitamiento general de la transparencia aumenta la probabilidad de que los usuarios no puedan comprender por qué se han eliminado contenidos determinados en ausencia de notificaciones adaptadas a los casos concretos.

38. *Apelaciones y reparación.* Las plataformas admiten la apelación de una serie de medidas, desde la eliminación de un perfil o una página a la eliminación de mensajes, fotografías o vídeos concretos[113]. No obstante, aun cuando se permita la apelación, las reparaciones a que pueden aspirar los usuarios parecen limitadas o extemporáneas hasta el punto de la no existencia y, en cualquier caso, resultan opacas para la mayoría de los usuarios e incluso expertos de la sociedad civil. Puede ser, por ejemplo, que la restitución de contenido sea una respuesta insuficiente en caso de que su eliminación hubiese dado lugar a un perjuicio específico —que puede ser físico, moral, financiero o para la reputación— para la persona autora del mensaje. Del mismo modo, la suspensión de una cuenta o la eliminación de un contenido mientras se produce una protesta o debate públicos podría tener importantes repercusiones en los derechos políticos y, sin embargo, la empresa no puede ofrecer una reparación por ello.

[109] Cf. Wikipedia: BOLD, revert, discuss cycle. Reddit alienta a sus moderadores que ofrezcan "a los usuarios nuevos o confusos explicaciones de las normas, consejos y enlaces que le sirvan de ayuda" (Reddit Moddiquette).

[110] Políticas de YouTube (políticas sobre el desnudo o el contenido sexual). Ayuda de YouTube, "Influencia de los creadores en YouTube".

[111] Comunicaciones de ARTICLE 19, pág. 7, y la Asociación para el Progreso de las Comunicaciones, pág. 16.

[112] Véase https://twitter. Com/TwitterSafety/status/971882517698510848 /.

[113] Fundación de la Frontera Electrónica y Visualizing Impact, "How to appeal", onlinecensorship.org. Facebook e Instagram únicamente permiten apelar contra la suspensión de una cuenta. Cf. Comunicación de Github, pág. 6.

Transparencia

39. Las empresas han elaborado informes de transparencia en los que publican datos agregados sobre las solicitudes de retirada de contenidos y datos de los usuarios. En esos informes se da fe del tipo de presiones a que se enfrentan las empresas. En los informes de transparencia se identifica, país por país, el número de solicitudes de eliminación de carácter legal[114], el número de solicitudes que dieron lugar a la adopción de algún tipo de medida o a la restricción de contenido[115] y, cada vez con mayor frecuencia, la descripción y la cita de ejemplos de determinados fundamentos jurídicos[116].

40. Sin embargo, como se afirma en el más importante examen de la transparencia de Internet realizado hasta la fecha, las empresas revelan "la menor cantidad de información posible sobre la manera en que se elaboran y aplican las normas y los mecanismos *privados* para la regulación por sí solas o de manera conjunta"[117]. En particular, el nivel de divulgación de información sobre las medidas adoptadas en cumplimiento de las solicitudes de retirada de origen privado con arreglo a las condiciones de servicio es "increíblemente bajo"[118]. Los criterios sobre el contenido están redactados en términos generales, lo que deja a la plataforma un amplio margen de discrecionalidad sobre lo que las empresas no informan suficientemente. El escrutinio público y de los medios de comunicación ha llevado a las empresas a complementar las políticas generales con mensajes explicativos[119] y algunos ejemplos hipotéticos[120] que, no obstante, quedan lejos de explicar los detalles de cómo se elaboran y aplican las normas internas[121]. Si bien las condiciones de servicio se publican generalmente en los idiomas locales, no sucede lo mismo con los informes de transparencia, las páginas informativas de la empresa y contenidos conexos, con lo que los usuarios que no son de habla inglesa reciben aún menos aclaraciones. En consecuencia, los usuarios, las autoridades públicas y la sociedad civil a menudo expresan su insatisfacción acerca de la imprevisibilidad de las actuaciones relacionadas con el servicio[122]. La falta de una participación suficiente, junto con el aumento de las críticas del público, ha obligado a las empresas a mantenerse en un constante estado de evaluación, revisión y defensa de las normas.

IV. Principios de derechos humanos para la moderación del contenido por parte de las empresas

41. El fundador de Facebook expresó recientemente su esperanza en poder elaborar un proceso en el que la empresa "pudiera reflejar con más exactitud los valores de la comunidad en distintos lugares"[123]. Ese proceso, y las normas pertinentes, pueden encontrarse en el derecho de los derechos humanos. Las normas privadas, que varían en función del modelo de negocio de cada empresa y las vagas afirmaciones acerca de los intereses de la comunidad, han creado un entorno inestable, imprevisible e inseguro para los

[114] Twitter Transparency Report: Removal Requests (enero a junio de 2017); Google Transparency Report: Government Requests to Remove Content; 2016 Reddit Inc., Transparency Report. Facebook no comunica el número total de solicitudes recibidas en cada país.

[115] Véase, por ejemplo, Facebook Transparency Report (Francia) (enero a junio de 2017); Google Transparency Report: Government Requests to Remove Content (India); Twitter Transparency Report (Turquía).

[116] *Ibid.*

[117] Comunicación de Ranking Digital Rights, pág. 4. En cursiva en el original.

[118] *Ibid.*, pág. 10.

[119] Véase Elliot Schrage, "Introducing hard questions", Facebook Newsroom, 15 de junio de 2017; Seguridad de Twitter, "La aplicación de nuevas normas para reducir la conducta inaceptable y el comportamiento abusivo", 18 de diciembre de 2017.

[120] Véanse, por ejemplo, las políticas de YouTube (políticas sobre el contenido violento o explícito).

[121] Angwin y Grasseger, "Facebook's secret censorship rules".

[122] Comunicaciones de Ranking Digital Rights, pág. 10; OBSERVACOM, pág. 10; Asociación para el Progreso de las Comunicaciones, pág. 17; Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas, págs. 4 y 5; Access Now, pág. 17; y EDRi, pág. 5.

[123] Kara Swisher y Kurt Wagner, "Here's the transcript of Recode's interview with Facebook CEO Mark Zuckerberg about the Cambridge Analytica controversy and more", *Recode*, 22 de marzo de 2018.

usuarios y han contribuido a intensificar la vigilancia de los Gobiernos. Las leyes nacionales no son adecuadas para las empresas que buscan normas comunes para una base de usuarios que es geográfica y culturalmente diversa. Sin embargo, las normas de derechos humanos, si se aplican de manera transparente y coherente, con aportaciones pertinentes de la sociedad civil y los usuarios, proporcionan un marco para la responsabilidad de los Estados y las empresas ante los usuarios a través de las fronteras nacionales.

42. Un marco de derechos humanos facilita la aplicación de respuestas normativas frente a las restricciones estatales excesivas —siempre que las empresas se rijan por normas similares. Los Principios Rectores y las "normas no vinculantes" que los acompañan proporcionan orientación sobre la forma en que las empresas deben evitar o mitigar las exigencias de los Gobiernos de una retirada de contenido excesiva. En ellos, eso sí, también se establecen principios de diligencia debida, transparencia, rendición de cuentas y reparación que limitan la interferencia de las plataformas con los derechos humanos con la elaboración de políticas y el desarrollo de productos. Las empresas comprometidas con la aplicación de las normas de derechos humanos en todas sus actividades —y no simplemente cuando conviene a sus intereses— se encontrarán en un terreno más firme cuando traten de hacer que los Estados cumplan las mismas normas. Por otra parte, cuando las empresas armonicen más estrechamente sus condiciones de servicio con el derecho de los derechos humanos, a los Estados les resultará más difícil sacar partido de ellas para censurar el contenido.

43. Los principios de derechos humanos también permiten a las empresas crear un entorno incluyente que dé cabida a las diversas necesidades e intereses de sus usuarios, estableciendo al mismo tiempo unas normas básicas de comportamiento previsibles y coherentes. En medio de un creciente debate sobre si las empresas ejercen una combinación de funciones de intermediación y edición, el derecho de los derechos humanos ofrece a los usuarios la promesa de que pueden contar con unas normas fundamentales para proteger su libertad de expresión más allá de lo que la legislación nacional podría limitar[124]. Sin embargo, el derecho de los derechos humanos no es tan inflexible o dogmático como para obligar a las empresas a permitir expresiones que puedan menoscabar los derechos de los demás o la capacidad de los Estados para proteger los legítimos intereses relacionados con la seguridad nacional o el orden público. En relación con una serie de comportamientos nocivos que quizá puedan tener mayor impacto en el espacio digital que fuera de él —como el acoso homófobo o misógino orientado a silenciar a las mujeres y las minorías sexuales, o la incitación a la violencia de todo tipo— el derecho de los derechos humanos no privaría de instrumentos a las empresas. Al contrario, ofrecería un marco mundialmente reconocido para la elaboración de esos instrumentos y un vocabulario común para explicar a los usuarios y a los Estados su naturaleza, finalidad y aplicación.

A. Normas sustantivas para la moderación del contenido

44. La era digital facilita una rápida difusión y un enorme alcance del contenido, pero también carece de los matices del contexto humano. Según los Principios Rectores, las empresas pueden tener en cuenta el tamaño, la estructura y las funciones distintivas de sus plataformas al evaluar la necesidad y la proporcionalidad de las restricciones del contenido.

45. *Los derechos humanos por defecto*. Las condiciones de servicio deberían abandonar el enfoque discrecional anclado en unas necesidades "comunitarias" genéricas y centradas en sí mismas. En vez de eso, las empresas deberían asumir compromisos de políticas de alto nivel para mantener unas plataformas en que los usuarios puedan desarrollar sus opiniones, expresarse libremente y acceder a información de todo tipo de una manera coherente con las normas de derechos humanos[125]. Esos compromisos deberían orientar su enfoque de la moderación del contenido y de problemas complejos como la propaganda

[124] Comunicación de Global Partners Digital, pág. 3; Principios Rectores, principio 11.
[125] Principios Rectores, principio 16.

computacional[126] y la recopilación y manipulación de datos de los usuarios. Las empresas deberían incorporar directamente en sus condiciones de servicio y sus "normas comunitarias" los principios pertinentes del derecho de los derechos humanos que garanticen que las medidas relacionadas con el contenido se guiarán por las mismas normas de legalidad, necesidad y legitimidad que rigen la regulación de la expresión por los Estados[127].

46. *"Legalidad".* Las normas de las empresas normalmente carecen de la claridad y concreción que permitiría a los usuarios predecir con una certeza razonable qué contenidos los colocan en el lado peligroso de la línea. Eso es especialmente evidente en el contexto del discurso "extremista" y de odio, esferas de restricción fácilmente susceptibles de acarrear retiradas excesivas de contenidos en ausencia de una evaluación rigurosa del contexto por una persona. Para complicar aún más el entendimiento por parte del público de unas normas específicas según el contexto aparece la nueva excepción general relativa al "carácter noticioso"[128]. Si bien se ve con agrado el reconocimiento del interés público, las empresas deberían también explicar qué factores se valoran a la hora de determinar ese interés público y qué otros factores, se tienen en cuenta al evaluar el carácter noticioso. Las empresas deberían hacer más por explicar sus normas de forma más detallada con datos agregados que ilustren las tendencias en el ámbito de la vigilancia del cumplimiento de las normas y ejemplos de casos reales o casos hipotéticos detallados que aclaren los matices de la interpretación y la aplicación de normas concretas.

47. *Necesidad y proporcionalidad*. Las empresas no solo deberían describir más detalladamente las normas controvertidas y aplicables en contextos específicos. También deberían divulgar datos y ejemplos que permitan conocer mejor los factores que tienen en cuenta a la hora de determinar una infracción, su gravedad y las medidas adoptadas como respuesta. En el contexto del discurso de odio, explicar cómo se resuelven algunos casos concretos puede ayudar a los usuarios a entender mejor la forma en que las empresas enfocan la difícil distinción entre el contenido ofensivo y la incitación al odio, o de qué forma se tienen en cuenta consideraciones como la intención del autor o la probabilidad de generar violencia en los contextos en línea. Los datos sobre las medidas adoptadas también servirían para establecer una base sobre la cual evaluar en qué medidas las compañías aplican las restricciones de una forma muy estricta. Debería explicarse en qué circunstancias aplican restricciones menos estrictas (como avisos, restricciones de edad o desmonetización).

48. *No discriminación*. Unas garantías genuinas de no discriminación requerirían que las empresas fueran más allá de los enfoques formalistas que consideran todas las características protegidas como igualmente vulnerables a los abusos, el acoso y otras formas de censura[129]. De hecho, esos enfoques parecerían poco coherentes con el énfasis que ellas mismas ponen en los asuntos relacionados con el contexto. En vez de eso, cuando las empresas elaboren o modifiquen sus políticas o productos deberían tratar de conocer cuáles son las preocupaciones de las comunidades que históricamente se han enfrentado al peligro de la censura y la discriminación y tener en cuenta esas preocupaciones.

B. Procesos para la moderación y actividades conexas por parte de las empresas

Respuestas a las solicitudes de los Gobiernos

49. Como se muestra en los informes de transparencia de las empresas, los Gobiernos las presionan para que supriman contenidos, suspendan cuentas e identifiquen y divulguen

[126] Véase Samuel Wooley y Philip Howard, *Computational Propaganda Worldwide: Executive Summary* (documento de trabajo núm. 2017.11 del Computational Propaganda Research Project) (Oxford, 2017).

[127] Comunicación de Global Partners Digital, págs. 10 a 13.

[128] Véase Joel Kaplan, "Input from community and partners on our community standards", Facebook Newsroom, 21 de octubre de 2016; normas y políticas de Twitter.

[129] Véase, por ejemplo, Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial, arts. 1, párrafo 4; y 2, párrafo 2.

información sobre las cuentas. Cuando la legislación local lo requiere, podría parecer que las empresas no tienen más opción que cumplirla. Sin embargo, las empresas pueden elaborar instrumentos que prevengan o mitiguen los riesgos para los derechos humanos causados por la legislación nacional o por exigencias incompatibles con las normas internacionales.

50. *Prevención y mitigación*. Las empresas a menudo afirman tomarse los derechos humanos muy en serio. No obstante, no basta con que las empresas contraigan esos compromisos a nivel interno y ofrezcan al público garantías *ad hoc* cuando surge la controversia. Las empresas también deberían, a los más altos niveles de dirección, adoptar y divulgar públicamente políticas específicas en las que "se instruya a todas las unidades de la empresa, incluidas las delegaciones locales, que resuelvan cualquier ambigüedad a favor del respeto a la libertad de expresión, la privacidad y demás derechos humanos". De esos compromisos deberían derivarse unas políticas y unos procedimientos en virtud de los cuales las exigencias de los Gobiernos se interpreten y apliquen de manera estricta y "garanticen que la restricción del contenido se reduzca a un nivel mínimo"[130]. Las empresas deben velar por que las solicitudes se formulen por escrito, contengan bases jurídicas concretas y válidas para las restricciones solicitadas y estén expedidas por una autoridad gubernamental facultada al efecto en un formato apropiado[131].

51. Cuando se encuentren ante solicitudes problemáticas, las empresas deben solicitar aclaraciones o modificaciones; recabar la asistencia de la sociedad civil, otras empresas afines, autoridades gubernamentales competentes, órganos regionales o internacionales y otros interesados; y explorar todas las opciones que la ley ofrezca para impugnarlas[132]. Cuando las empresas reciban de los Estados solicitudes formuladas al amparo de sus condiciones de servicio u otros medios extralegales, deberían canalizarlas a través de procedimientos de observancia de la legalidad y evaluar la validez de dichas solicitudes en el marco de la legislación local y las normas de derechos humanos.

52. *Transparencia*. Enfrentados a la censura y los riesgos asociados para los derechos humanos, los usuarios solo pueden adoptar decisiones informadas sobre su participación en los medios sociales y la forma de desarrollarla si las interacciones de las empresas y los Estados son verdaderamente transparentes. Es necesario elaborar mejores prácticas sobre cómo ofrecer esa transparencia. La información de las empresas acerca de las solicitudes de los Estados debe complementarse con datos relativos a los tipos de solicitudes recibidas (por ejemplo, en relación con la difamación, el discurso de odio, el contenido relacionado con el terrorismo, etc.) y las medidas adoptadas (como la retirada parcial o total del contenido, la retirada a nivel mundial o en un país determinado, la suspensión de la cuenta, la retirada contemplada en las condiciones de servicio, etc.). Las empresas también deben proporcionar ejemplos específicos con la mayor frecuencia posible[133]. La información sobre la transparencia debería ampliarse a las solicitudes de los Gobiernos formuladas al amparo de las condiciones de servicio de las empresas[134] y deben tenerse en cuenta también las iniciativas público-privadas para restringir el contenido, como el Código de Conducta de la Unión Europea para contrarrestar el discurso de odio ilegal en línea, las iniciativas gubernamentales como las unidades de referencia de Internet y los entendimientos bilaterales como los que se han comunicado entre YouTube y el Pakistán o entre Facebook e Israel. Las empresas deben conservar registros de las solicitudes formuladas en el marco de esas iniciativas y las comunicaciones entre la empresa y el solicitante, y estudiar la forma de remitir copias de dichas solicitudes a archivos de terceros.

[130] Véase A/HRC/35/22, párrs. 66 y 67.

[131] Comunicaciones de Global Network Initiative, págs. 3 y 4 y GitHub, págs. 3 a 5.

[132] Véase A/HRC/35/22, párr. 68.

[133] Véase, por ejemplo, Twitter Transparency Report: Removal Requests (enero a junio de 2017).

[134] Twitter ha comenzado a publicar datos sobre "solicitudes extralegales presentadas por representantes conocidos de Gobiernos relativas a contenido que pudiera vulnerar las normas de Twitter en virtud de las cuales se prohíbe el comportamiento abusivo, la promoción del terrorismo y la vulneración de la propiedad intelectual. *Ibid.* Véase también Microsoft, Content Removal Requests Report (enero a junio de 2017).

Formulación de normas y desarrollo de productos

53. *Diligencia debida*. Aunque varias empresas se comprometen con el respeto de la diligencia debida en materia de derechos humanos al evaluar su respuesta a las restricciones impuestas por el Estado, no está claro si aplican las mismas salvaguardias a la hora de prevenir o mitigar los riesgos para la libertad de expresión que plantean el desarrollo y la aplicación de sus propias políticas[135]. Las empresas deberían elaborar criterios claros y precisos para identificar las actividades que dan lugar a dichas evaluaciones. Además de la revisión de las políticas y los procesos de moderación del contenido, deberían realizarse evaluaciones sobre la vigilancia de los canales de los usuarios y otras formas de difusión de contenidos, la introducción de nuevas características o servicios y la modificación de los existentes, el desarrollo de tecnologías de automatización y las decisiones relativas a la entrada en el mercado, como los acuerdos para proporcionar versiones de la plataforma adaptadas a un país concreto[136]. En los informes elaborados en el pasado se especificaban también las cuestiones que debían tenerse en cuenta en esas evaluaciones y los procesos internos y la formación necesarios para incorporar las evaluaciones y sus conclusiones en las operaciones pertinentes. Además, esas evaluaciones deben ser de carácter continuo y adaptarse a los cambios de las circunstancias o del contexto operativo[137]. Las iniciativas de múltiples interesados, como la Global Network Initiative ofrecen una vía para que las empresas elaboren y perfeccionen las mencionadas evaluaciones y otros procesos relacionados con la diligencia debida.

54. *Aportación y participación del público*. Quienes participaron en las consultas plantearon habitualmente la preocupación de que las empresas no mantuvieran un contacto apropiado con los usuarios y la sociedad civil, especialmente en el Sur global. La aportación de los titulares de derechos afectados (o sus representantes) y los expertos locales sobre la cuestión, y unos procesos internos de adopción de decisiones que incorporen de una forma satisfactoria las observaciones recibidas son componentes intrínsecos de la diligencia debida[138]. Las consultas —especialmente las planteadas de forma general, como la solicitud de comentarios del público— permiten a las empresas analizar el impacto de sus actividades en los derechos humanos desde diversas perspectivas, al tiempo que las alientan a prestar más atención a la forma en que unas normas aparentemente benignas o claramente "favorables a la comunidad" pueden tener efectos importantes, "hiperlocales" en las comunidades[139]. Por ejemplo, la interacción con una variedad de grupos indígenas diversa desde el punto de vista geográfico puede ayudar a las empresas a elaborar mejores indicadores para tener en cuenta el contexto artístico y cultural a la hora de evaluar el contenido en el que aparezcan desnudos.

55. *Transparencia en la elaboración de normas*. Con demasiada frecuencia, las empresas parecen introducir modificaciones en sus normas y sus productos sin atenerse a la debida diligencia en materia de derechos humanos o sin evaluar el impacto en casos reales. Las empresas deberían, cuando menos, recabar de los expertos y los usuarios interesados comentarios sobre sus evaluaciones de los impactos de formas que garanticen, de ser necesario, la confidencialidad de dichos comentarios. También deberían comunicar al público de manera clara los procesos y normas que dieron lugar a dichas evaluaciones.

Observancia de las normas

56. *Automatización y evaluación por las personas*. La moderación automatizada del contenido, derivada del alcance y la escala enorme del contenido generado por los usuarios, plantea riesgos específicos de que se adopten medidas relacionadas con el contenido que sean incompatibles con las normas de derechos humanos. En la responsabilidad de las

[135] Comunicación de Ranking Digital Rights, pág. 12; Principios Rectores, principio 17.

[136] Véase A/HRC/35/22, párr. 53.

[137] *Ibid*., párrs. 54 a 58.

[138] Véanse Principios Rectores, principio 18; y A/HRC/35/22, párr. 57.

[139] Chinmayi Arun, "Rebalancing regulation of speech: hyper-local content on global web-based platforms", Berkman Klein Center for Internet and Society Medium Collection, Universidad de Harvard, 2018; *Pretoria News*, "Protest at Google, Facebook 'bullying' of bare-breasted maidens", 14 de diciembre de 2017.

empresas de prevenir y mitigar los efectos en los derechos humanos deberían tenerse en cuenta las importantes limitaciones de que adolece la automatización, como las dificultades para tener en cuenta el contexto, la amplia variación de matices idiomáticos y el significado y las particularidades lingüísticas y culturales. La automatización derivada de entendimientos alcanzados con el país en que está establecida la empresa puede acarrear una grave discriminación entre usuarios de diversas partes del mundo. Como mínimo, la tecnología desarrollada para abordar aspectos relacionados con la magnitud de la escala debería estar sometida a una auditoría rigurosa y contar con aportaciones de los usuarios y de la sociedad civil.

57. La responsabilidad de promover unas prácticas de moderación del contenido de gran precisión y sensibles al contexto que respeten la libertad de expresión también requiere que las empresas fortalezcan y garanticen la profesionalización de las personas que se dedican a evaluar los contenidos señalados. Ese fortalecimiento debe incluir una protección para los moderadores que sea compatible con las normas de derechos humanos aplicables a los derechos laborales y un compromiso profundo de incorporar conocimientos culturales, lingüísticos y de otro tipo en cada uno de los mercados en que operen. También deberían diversificarse la dirección de la empresa y los equipos de política para facilitar que en la evaluación del contenido intervengan los expertos locales.

58. *Notificación y apelación.* Tanto los usuarios como los expertos de la sociedad civil suelen mostrar preocupación por la limitada información de que disponen quienes sufren de la eliminación de su contenido o la suspensión o desactivación de su cuenta, o quienes informan de abusos como el acoso misógino o la revelación de documentos privados y confidenciales. La falta de información contribuye a crear un entorno de normas secretas, incompatibles con las normas de claridad, especificidad y predecibilidad. Eso interfiere con la capacidad de las personas para impugnar las medidas adoptadas con respecto al contenido o hacer un seguimiento de las quejas relacionadas con el contenido. En la práctica, sin embargo, la ausencia de unos mecanismos sólidos para apelar la eliminación de contenidos favorece a quienes los señalan por encima de quienes los publican. Algunos dirán que consumiría mucho tiempo y recursos permitir que se pudiera apelar cualquier medida adoptada en relación con los contenidos. Sin embargo, las empresas podrían colaborar entre sí y con la sociedad civil para analizar soluciones que pudieran extenderse después, como los programas de *ombudsman* bien de compañías concretas bien a nivel de todo el sector. Entre las mejores ideas sobre esos programas cabe mencionar un "consejo de los medios sociales" independiente, creado a imagen de los consejos de prensa que facilitan mecanismos de queja a nivel de todo el sector y promueven la reparación de las injusticias[140]. Ese mecanismo podría recibir las quejas de los usuarios que reúnan determinados criterios y recoger los comentarios del público con respecto a problemas recurrentes relacionados con la moderación del contenido, como la aplicación de una censura excesiva en un determinado tema. Los Estados deberían prestar su apoyo a los mecanismos de apelación que puedan extenderse y que operen de manera compatible con las normas de derechos humanos.

59. *Reparación*. En los Principios Rectores se pone de relieve la responsabilidad de reparar "los efectos adversos" (principio 22). No obstante, son pocas las empresas, si es que hay alguna, que contemplan la reparación. Las empresas deberían instituir programas sólidos de reparación, que podrían ir de la readmisión y el reconocimiento de los errores hasta acuerdos relacionados con el daño ocasionado a la reputación u otros tipos de daño. Se ha producido cierta convergencia entre varias empresas en cuanto al contenido de sus normas, dando lugar a la posibilidad de que las empresas colaboren para ofrecer una reparación por conducto de un consejo de los medios sociales, otros programas de *ombudsman* o la decisión arbitral de un tercero. Si aun así no se consigue llegar a la reparación, puede ser necesaria la intervención legislativa o judicial.

60. *Autonomía de los usuarios*. Las empresas han elaborado instrumentos que permiten a los usuarios configurar sus propios entornos en línea. Eso incluye silenciar o bloquear a otros usuarios o determinados tipos de contenido. Del mismo modo, las plataformas suelen

[140] Véase ARTICLE 19, *Self-regulation and 'Hate Speech' on Social Media Platforms* (Londres, 2018), págs. 20 a 22.

permitir que los usuarios creen grupos cerrados o privados, sujetos a moderación por sus propios miembros. Si bien las normas sobre el contenido que se aplican en los grupos cerrados deben ser compatibles con las normas básicas de derechos humanos, las plataformas deberían alentar la creación de esos grupos afines, dada su utilidad en lo que se refiere a proteger la libertad de opinión, ampliar el espacio para las comunidades vulnerables y poner a prueba ideas controvertidas o impopulares. Debería huirse del requisito de revelar la identidad real, dadas las repercusiones negativas en cuanto a la privacidad y la seguridad de las personas vulnerables[141].

61. El aumento de la preocupación acerca de la verificabilidad, relevancia y utilidad de la información en línea plantea complejas cuestiones acerca de cómo las empresas deben respetar el derecho de acceso a la información. Como mínimo, las empresas deben desvelar los detalles relativos a sus enfoques de la moderación. Si las empresas clasifican el contenido de los canales de los medios sociales sobre la base de las interacciones entre los usuarios, deberían explicar qué datos se recogen acerca de dichas interacciones y cómo se tienen en cuenta en los criterios de clasificación. Las empresas deben proporcionar a todos los usuarios la posibilidad real de desvincularse del tratamiento realizado por la plataforma[142].

Transparencia de la adopción de decisiones

62. A pesar de los adelantos en la transparencia global de las solicitudes de supresión de contenidos presentadas por los Gobiernos, las medidas amparadas en las condiciones de servicio normalmente no se comunican. Las empresas no publican datos sobre el volumen y el tipo de las solicitudes privadas que reciben en el marco de esas condiciones, y menos aún sobre las tasas de cumplimiento de dichas solicitudes. Las empresas deben poner en marcha iniciativas de fomento de la transparencia que sirvan para explicar los efectos de la automatización, la moderación realizada por personas y la detección de contenido problemático por los usuarios o la "detección fiable" en las medidas relacionadas con las condiciones de servicio. Aunque algunas empresas comienzan a facilitar cierta información sobre esas medidas, el sector debería avanzar para proporcionar más detalles sobre casos específicos y representativos y avances importantes en la interpretación y la aplicación de sus políticas.

63. Las empresas aplican la "ley de la plataforma", adoptando medidas en aspectos relacionados con el contenido sin divulgar realmente la motivación de dichas medidas. En una situación ideal, las empresas deberían elaborar una especie de jurisprudencia que permita a los usuarios, a la sociedad civil y a los Estados comprender la forma en que las empresas interpretan y aplican sus normas. Si bien ese sistema de "jurisprudencia" no supondría mantener el nivel de presentación de informes que el público espera de los tribunales y órganos administrativos, un archivo detallado de casos y ejemplos aclararía las normas en una medida similar a como lo hace la información sobre los casos[143]. Un consejo de los medios sociales facultado para evaluar las quejas en todo el sector de las TIC podría ser el mecanismo creíble e independiente encargado de aportar esa transparencia.

V. Recomendaciones

64. **Fuerzas opacas están conformando la capacidad de personas de todo el mundo de ejercer su libertad de expresión. El momento actual exige una mejora radical de la transparencia, una rendición de cuentas genuina y un verdadero compromiso con la reparación con el fin de proteger la capacidad de las personas para utilizar las plataformas en línea como foros donde expresarse libremente, acceder a la**

[141] Véase el párrafo 30.

[142] Facebook, por ejemplo, permite a los usuarios acceder a las historias contenidas en su News Feed en orden cronológico inverso, pero advierte de que "finalmente" volverá a sus condiciones de edición habituales. Ayuda de Facebook, "What's the difference between top stories and most recent stories on News Feed?".

[143] Véase, por ejemplo, Madeleine Varner y otros, "What does Facebook consider hate speech?", *ProPublica*, 28 de diciembre de 2017.

información y participar en la vida pública. En el presente informe se han identificado varias medidas, incluidas las que figuran a continuación.

Recomendaciones dirigidas a los Estados

65. **Los Estados deberían derogar cualquier ley en virtud de la cual se penalice o se restrinja indebidamente la libertad de expresión, tanto en línea como fuera de línea.**

66. **Una normativa inteligente, y no una excesivamente estricta con una única óptica, debería ser la norma, una normativa centrada en garantizar la transparencia de las empresas y en la reparación para que el público pueda elegir si desea participar en foros en línea y cómo quiere hacerlo. Los Estados solo deberían tratar de restringir los contenidos por medio de una orden dictada por una autoridad judicial imparcial e independiente, y de acuerdo con las debidas garantías procesales y las normas establecidas de legalidad, necesidad y legitimidad. Los Estados deberían abstenerse de imponer sanciones desproporcionadas, ya se trate de cuantiosas multas o de penas de prisión, a los intermediarios de Internet, dado su importante efecto paralizante sobre la libertad de expresión.**

67. **Los Estados y las organizaciones intergubernamentales deberían abstenerse de promulgar leyes o concertar acuerdos que requieran la vigilancia o el filtrado "activo" del contenido, que es incompatible con el derecho a la privacidad y puede ser equivalente a la censura previa a la publicación.**

68. **Los Estados deberían abstenerse de adoptar modelos de regulación en los que sean los organismos gubernamentales, y no las autoridades judiciales, quienes se erijan en los árbitros de lo que es una expresión legítima. Deberían evitar delegar en las empresas la responsabilidad como evaluadores del contenido, con lo que se pone el juicio de las empresas por encima de los valores de derechos humanos, en detrimento de los usuarios.**

69. **Los Estados deberían publicar informes de transparencia detallados sobre todas las solicitudes relacionadas con el contenido presentadas a los intermediarios y recabar de una forma genuina la participación del público en todas las consideraciones normativas.**

Recomendaciones dirigidas a las empresas de TIC

70. **Las empresas deberían reconocer que el marco autorizado a nivel mundial para garantizar la libertad de expresión en sus plataformas es el derecho de los derechos humanos, no las diversas leyes de los Estados o sus propios intereses privados, y deberían volver a examinar en consecuencia sus normas sobre el contenido. El derecho de los derechos humanos proporciona a las empresas los instrumentos necesarios para diseñar y elaborar políticas y procesos que respeten las normas democráticas y rechacen las exigencias autoritarias. Ese enfoque comienza con unas normas basadas en los derechos, continúa con unas evaluaciones rigurosas del impacto del desarrollo de productos y la elaboración de políticas en los derechos humanos, y llega hasta las operaciones con una evaluación, una reevaluación y una consulta genuina con el público y con la sociedad civil de carácter continuo. Los Principios Rectores sobre las Empresas y los Derechos Humanos, junto con unas directrices específicas para el sector elaboradas por la sociedad civil, los órganos intergubernamentales, la Global Network Initiative y otros interesados, constituyen unos enfoques de referencia que todas las empresas de Internet deberían adoptar.**

71. **Las empresas deben adoptar enfoques de la transparencia radicalmente distintos en todas las etapas de sus operaciones, desde la elaboración de normas hasta la recopilación y aplicación de "jurisprudencia" que enmarque la interpretación de las normas privadas. La transparencia requiere una mayor interacción con las organizaciones de derechos digitales y otros sectores pertinentes de la sociedad civil y la renuncia a concertar acuerdos secretistas con los Estados sobre las normas relativas a los contenidos y su aplicación.**

72. **Teniendo en cuenta su impacto en la esfera pública, las empresas deben abrirse a la rendición de cuentas. Los consejos de prensa eficaces y respetuosos de los derechos que funcionan en todo el mundo pueden tomarse como modelo para imponer unos niveles mínimos de coherencia, transparencia y rendición de cuentas en el ámbito de la moderación del contenido por las empresas. Los enfoques no gubernamentales de terceros, si se basan en las normas de derechos humanos, podrían proporcionar ejemplos de mecanismos de apelación y reparación sin entrañar unos costos prohibitivos que disuadan a las empresas pequeñas o recién llegadas al mercado. Todos los segmentos del sector de las TIC que actúan como moderadores del contenido o como vigilantes deberían considerar como una prioridad fundamental el desarrollo de mecanismos de rendición de cuentas a nivel de todo el sector (como los consejos de los medios sociales).**