



# Asamblea General

Distr. general  
9 de octubre de 2019  
Español  
Original: inglés

---

## Septuagésimo cuarto período de sesiones

Tema 70 b) del programa

**Promoción y protección de los derechos humanos:  
cuestiones de derechos humanos, incluidos otros medios  
de mejorar el goce efectivo de los derechos humanos y las  
libertades fundamentales**

## **Promoción y protección del derecho a la libertad de opinión y de expresión\***

### **Nota del Secretario General**

El Secretario General tiene el honor de transmitir a la Asamblea General el informe preparado por el Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, David Kaye, presentado de conformidad con la resolución 34/18 del Consejo de Derechos Humanos. En el presente informe, el Relator Especial evalúa el derecho de los derechos humanos que se aplica a la reglamentación del “discurso de odio” en línea.

---

\* Este informe se presentó fuera del plazo establecido con el propósito de incluir la información más reciente.



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

### *Resumen*

En un mundo en el que cada vez se piden más límites al discurso de odio, el derecho internacional de los derechos humanos establece normas que rigen los enfoques de los Estados y las empresas respecto de la expresión en línea. En este informe, presentado de conformidad con la resolución 34/18 del Consejo de Derechos Humanos, el Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión explica cómo esas normas proporcionan un marco para que los Gobiernos examinen las opciones normativas y las empresas determinen cómo respetar los derechos humanos en línea. El Relator Especial comienza con una introducción al marco jurídico internacional, centrándose en los tratados de las Naciones Unidas y en las principales interpretaciones de las disposiciones relacionadas con lo que se denomina coloquialmente “discurso de odio”. Después destaca las principales obligaciones de los Estados y aborda la forma en que la moderación de los contenidos por las empresas puede respetar los derechos humanos de los usuarios y del público. Concluye con recomendaciones para los Estados y las empresas.

---

## Índice

	<i>Página</i>
I. Introducción. . . . .	4
II. La reglamentación del “discurso de odio” en el derecho internacional de los derechos humanos . . . . .	4
III. Reglamentación del discurso de odio en línea . . . . .	13
A. Obligaciones del Estado y reglamentación del discurso de odio en línea . . . . .	13
B. Moderación de los contenidos por las empresas y discurso de odio . . . . .	17
IV. Conclusiones y recomendaciones. . . . .	24

## I. Introducción

1. El “discurso de odio”, una frase breve que no define el derecho internacional convencional, tiene una doble ambigüedad. Su vaguedad, y la falta de consenso en torno a su significado, pueden utilizarse indebidamente para permitir que se infrinja una amplia gama de expresiones lícitas. Muchos gobiernos utilizan el “discurso de odio” de forma similar a la que utilizan las “noticias falsas”, para atacar a los enemigos políticos, a los no creyentes, a los disidentes y a los críticos. Sin embargo, la debilidad de la frase (“son solo palabras”) también parece hacer que los gobiernos y las empresas no aborden daños genuinos, como los resultantes de discursos que incitan a la violencia o la discriminación contra los vulnerables o al silenciamiento de los marginados. La situación frustra a un público que a menudo percibe un abuso desenfrenado en línea.

2. En un mundo en el que cada vez se piden más límites al discurso de odio, el derecho internacional de los derechos humanos establece normas que rigen los enfoques de los Estados y las empresas respecto de la expresión en línea (A/HRC/38/35, para. 45)<sup>1</sup>. En el presente informe, el Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión explica cómo esas normas proporcionan un marco para que los gobiernos examinen las opciones normativas y las empresas determinen cómo respetar los derechos humanos en línea. El Relator Especial comienza con una introducción al marco jurídico internacional, centrándose en los tratados de las Naciones Unidas y en las principales interpretaciones de las disposiciones relacionadas con lo que se denomina coloquialmente “discurso de odio”. Después destaca las principales obligaciones de los Estados y aborda la forma en que la moderación de los contenidos por las empresas puede garantizar el respeto de los derechos humanos de los usuarios y del público. Concluye con recomendaciones para los Estados y las empresas.

3. El presente informe es el sexto de una serie de informes publicados desde 2015 en que el Relator Especial ha abordado las normas de derechos humanos aplicables a la libertad de opinión y de expresión en el sector de las tecnologías de la información y las comunicaciones (TIC)<sup>2</sup>. Debe leerse a la luz de las normas y recomendaciones propuestas anteriormente, que no se repiten necesariamente en el presente documento. Al igual que en sus informes anteriores, el Relator Especial se basa en gran medida en las normas internacionales existentes y en aportaciones considerables de la sociedad civil en los últimos años.

## II. La reglamentación del “discurso de odio” en el derecho internacional de los derechos humanos

4. Según el derecho internacional de los derechos humanos, la limitación del “discurso de odio” parece exigir que se reconcilien dos conjuntos de valores: los requisitos de la sociedad democrática de permitir un debate abierto y la autonomía y

---

<sup>1</sup> En el presente informe se utiliza el término “discurso de odio” para referirse a las obligaciones y limitaciones del derecho de los derechos humanos en que no se utiliza ese término en particular. Véase Susan Benesch, “Proposals for improved regulation of harmful online content”, documento preparado para el Israel Democracy Institute (2019). Benesch acuñó un término similar, “discurso peligroso”, para definir una “capacidad de catalizar la violencia de un grupo contra otro”. Véase también Susan Benesch, “Dangerous speech: a proposal to prevent group violence” (2012).

<sup>2</sup> Véanse A/HRC/29/32, sobre codificación y anonimato, A/HRC/32/38, sobre la recopilación de las repercusiones del sector de las TIC en los derechos, A/HRC/35/22, sobre la industria del acceso digital, A/HRC/38/35, sobre la moderación de contenidos en línea, y A/73/348, sobre inteligencia artificial y derechos humanos.

el desarrollo individuales, con la obligación, también imperiosa, de prevenir los ataques contra las comunidades vulnerables y garantizar la participación igualitaria y no discriminatoria de todas las personas en la vida pública<sup>3</sup>. A menudo los gobiernos se aprovechan de la incertidumbre resultante para amenazar expresiones legítimas, como la disidencia política y la crítica o el desacuerdo religioso<sup>4</sup>. Sin embargo, la libertad de expresión, los derechos a la igualdad y a la vida y la obligación de no discriminación se refuerzan mutuamente; el derecho de los derechos humanos permite a los Estados y a las empresas centrarse en la protección y promoción de la expresión de todos, especialmente de aquellos cuyos derechos suelen estar en peligro, al tiempo que aborda la discriminación pública y privada que socava el disfrute de todos los derechos.

### Libertad de expresión

5. El artículo 19 1) del Pacto Internacional de Derechos Civiles y Políticos protege el derecho a tener opiniones sin ser molestados, y el artículo 19 2) garantiza el derecho a la libertad de expresión, es decir, el derecho a buscar, recibir y difundir informaciones e ideas de toda índole, sin limitación de fronteras, por cualquier medio. Muchos otros tratados, mundiales y regionales, protegen expresamente la libertad de expresión<sup>5</sup>. El Comité de Derechos Humanos, órgano de supervisión del cumplimiento del Pacto, ha subrayado que estas libertades son “condiciones indispensables para el pleno desarrollo de la persona ... [y] constituyen la piedra angular de toda sociedad libre y democrática”. Forman “la base para el pleno goce de una amplia gama de otros derechos humanos”<sup>6</sup>.

6. Dado que la libertad de expresión es fundamental para el disfrute de todos los derechos humanos, las restricciones deben ser excepcionales, estar sujetas a condiciones rigurosas y a una supervisión estricta. El Comité de Derechos Humanos ha subrayado que las restricciones, incluso cuando estén justificadas, “no deberán poner en peligro ese derecho en sí mismo”<sup>7</sup>. El carácter excepcional de las limitaciones se describe en el artículo 19 3) del Pacto, reconociendo que los Estados pueden restringir la expresión en virtud del artículo 19 2) solo cuando lo disponga la ley y sea necesario para proteger los derechos o la reputación de los demás, la seguridad nacional o el orden público, o la salud o la moral públicas. Se trata de excepciones estrictamente definidas (véanse, en particular, [A/67/357](#), párr. 41, y [A/HRC/29/32](#), párrs. 32 a 35), y la justificación de la restricción debe hacerla la autoridad que restringe la libertad de expresión, no los oradores, que no tienen que

<sup>3</sup> Véase, en particular, el informe sobre el discurso de odio del anterior Relator Especial, Frank La Rue ([A/67/357](#)).

<sup>4</sup> *Ibid.*, párrs. 51 a 54.

<sup>5</sup> Véase, por ejemplo, la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial, art. 5; la Convención sobre los Derechos del Niño, art. 13; la Convención sobre los Derechos de las Personas con Discapacidad, art. 21; la Convención Internacional sobre la Protección de los Derechos de Todos los Trabajadores Migratorios y de Sus Familiares, art. 13; la Convención Americana sobre Derechos Humanos, art. 13; la Carta Africana de Derechos Humanos y de los Pueblos, art. 9; y el Convenio Europeo de Derechos Humanos, art. 10.

<sup>6</sup> Comité de Derechos Humanos, observación general núm. 34 (2011) relativa a la libertad de opinión y de expresión, párrs. 2 y 4; véase también *ibid.*, párrs. 5 y 6.

<sup>7</sup> *Ibid.*, párr. 21. El Comité de Derechos Humanos aclaró que “las restricciones no deben comprometer la esencia del derecho”, y añadió que “Las leyes que autoricen la aplicación de restricciones deben utilizar criterios precisos y no conferir una discrecionalidad sin trabas a los encargados de su aplicación” (Comité de Derechos Humanos, observación general núm. 27 (1999) sobre la libertad de circulación, párr. 13).

demostrar que tienen derecho a ella<sup>8</sup>. Cualquier limitación debe cumplir tres condiciones:

a) **Legalidad.** La restricción debe ser establecida por leyes que sean precisas, públicas y transparentes; debe evitar proporcionar a las autoridades una discreción ilimitada, y se debe notificar apropiadamente a aquellos cuyo discurso está siendo regulado. Las reglas deben estar sujetas a comentarios públicos y a un proceso legislativo o administrativo periódico. Las salvaguardias procesales, especialmente las garantizadas por tribunales o cortes independientes, deben proteger los derechos;

b) **Legitimidad.** La restricción debe estar justificada para proteger uno o más de los intereses especificados en el artículo 19 3) del Pacto, es decir, para respetar los derechos o la reputación de los demás o para proteger la seguridad nacional, el orden público, la salud o la moral públicas;

c) **Necesidad y proporcionalidad.** El Estado debe demostrar que la restricción es necesaria para proteger un interés legítimo y que es el medio menos restrictivo para lograr el objetivo perseguido. El Comité de Derechos Humanos se ha referido a esas restricciones como “pruebas estrictas”, según las cuales las “restricciones solamente se podrán aplicar para los fines con que fueron prescritas y deberán estar relacionadas directamente con la necesidad específica de la que dependen”<sup>9</sup>.

7. Los Estados hacen valer regularmente propósitos apropiados para imponer limitaciones de la expresión, pero no demuestran que sus limitaciones cumplen los criterios de legalidad o de necesidad y proporcionalidad (véase [A/71/373](#)). Por esta razón, las reglas deben aplicarse estrictamente y de buena fe, con una supervisión sólida y transparente. De conformidad con el artículo 2 3) b) del Pacto, los Estados están obligados a garantizar que las personas que deseen interponer un recurso por violación del Pacto tengan ese derecho determinado por “la autoridad competente, judicial, administrativa o legislativa, o cualquiera otra autoridad competente prevista por el sistema legal del Estado” (véase también [A/HRC/22/17/Add.4](#), párr. 31).

### **Apología del odio que constituye incitación**

8. De conformidad con el artículo 20 2) del Pacto, los Estados partes están obligados a prohibir por ley “toda apología del odio nacional, racial o religioso que constituya incitación a la discriminación, la hostilidad o la violencia”. Los Estados no están obligados a tipificar como delito ese tipo de expresiones. El anterior Relator Especial explicó que el artículo 20 2) se refiere a a) la apología del odio, b) la apología que constituye incitación, y c) la incitación que pueda llevar a discriminación, hostilidad o violencia ([A/67/357](#), para. 43).

9. Las normas de derechos humanos de las Naciones Unidas ofrecen una protección más amplia contra la discriminación que la que se presta mediante la atención que presta el artículo 20 2) al odio nacional, racial o religioso. El artículo 2 1) del Pacto garantiza los derechos de todas las personas “sin distinción alguna”, y el artículo 26 dispone expresamente que “la ley prohibirá toda discriminación y garantizará a todas las personas protección igual y efectiva contra cualquier discriminación”. Las normas internacionales garantizan la protección contra acciones adversas por motivos de raza, color, sexo, idioma, religión, opinión política o de otra índole, origen nacional o social, posición económica, nacimiento u otra condición, incluido el origen o identidad indígena, la discapacidad, la condición de migrante o refugiado, la orientación sexual, la identidad de género o la condición de

<sup>8</sup> Comité de Derechos Humanos, observación general núm. 34 (2011), párr. 27.

<sup>9</sup> *Ibid.*, párr. 22.

intersexual<sup>10</sup>. El ámbito de la protección se ha ido ampliando con el tiempo, de modo que otras categorías, como la edad o el albinismo, también gozan de protecciones explícitas hoy en día. Dada la expansión de la protección en todo el mundo, debe entenderse que la prohibición de la incitación se aplica a las categorías más amplias que abarca actualmente el derecho internacional de los derechos humanos.

10. Un punto crítico es que la persona cuya expresión va a ser prohibida en virtud del artículo 20 2) del Pacto es el apologeta cuya apología constituye incitación. En virtud del artículo 20 (o de cualquier otra disposición del derecho de los derechos humanos), no se debe silenciar a una persona que no haga apología del odio que constituya incitación a la discriminación, la hostilidad o la violencia, por ejemplo, una persona que abogue en pro de una minoría o incluso de una interpretación ofensiva de un principio religioso o un acontecimiento histórico, o una persona que comparta ejemplos de odio e incitación para informar o concienciar al respecto. Esa expresión debe ser protegida por el Estado, aun cuando este no esté de acuerdo con ella o se sienta ofendido por ella<sup>11</sup>. En el derecho internacional de los derechos humanos no existe un “veto para los abucheadores”<sup>12</sup>.

11. En la Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial, adoptada un año antes que el Pacto Internacional de Derechos Civiles y Políticos, se insta a los Estados a “erradicar toda incitación a la discriminación racial y todo acto de discriminación racial”, teniendo debidamente en cuenta otros derechos protegidos por el derecho de los derechos humanos, como la libertad de expresión (véanse los artículos 4 y 5 de la Convención). De conformidad con el artículo 4 de la Convención, los Estados partes están obligados, entre otras cosas, a: a) “declarar como acto punible conforme a la ley toda difusión de ideas basadas en la superioridad o en el odio racial, toda incitación a la discriminación racial, así como todo acto de violencia o toda incitación a cometer tales actos contra cualquier raza o grupo de personas de otro color u origen étnico”; y b) “declarar ilegales y prohibir las organizaciones, así como las actividades organizadas de propaganda y toda otra actividad de propaganda, que promuevan la discriminación racial e inciten a ella y reconocerán que la participación en tales organizaciones o actividades constituye un delito penado por la ley”.

12. El artículo 20 2) del Pacto Internacional de Derechos Civiles y Políticos y el artículo 4 de la Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial se refieren a categorías específicas de expresión, a menudo caracterizadas como “discurso de odio”<sup>13</sup>. La redacción de esas disposiciones es ambigua, en comparación con la del artículo 19 2) del Pacto<sup>14</sup>. Mientras que la libertad de expresión definida en el artículo 19 2) del Pacto implica derechos amplios encarnados en verbos activos (buscar, recibir, impartir) y el más amplio alcance posible (ideas de todo tipo, sin limitación de fronteras, por cualquier medio), las proscripciones en virtud del artículo 20 2) del Pacto y el artículo 4 de la Convención, aunque son mucho más específicas que las prohibiciones genéricas del discurso de odio, suponen un lenguaje de emociones difícil de definir (odio, hostilidad) y una prohibición muy específica del contexto (apología de la incitación). El Comité de

<sup>10</sup> Véase también Article 19, “*Hate Speech*” Explained: A Toolkit (Londres, 2015), pág. 14. Sobre la violencia contra la mujer en línea, véase [A/HRC/38/47](#).

<sup>11</sup> Comité de Derechos Humanos, observación general núm. 34 (2011), párr. 11.

<sup>12</sup> Véase Evelyn M. Aswad, “To ban or not to ban blasphemous videos”, *Georgetown Journal of International Law*, vol. 44, núm. 4 (2013).

<sup>13</sup> Véase Jeremy Waldron, *The Harm in Hate Speech* (Harvard University Press, 2012).

<sup>14</sup> La ambigüedad no es sorprendente, teniendo en cuenta la historia de la negociación. Véase Jacob Mchangama, “The sordid origin of hate-speech laws”, *Policy Review* (diciembre de 2011 y enero de 2012).

Derechos Humanos ha llegado a la conclusión de que los artículos 19 y 20 del Pacto “son compatibles entre sí y se complementan”<sup>15</sup>. Aun así, requieren interpretación.

13. En su observación general núm. 34 (2011), el Comité de Derechos Humanos consideró que cuando un Estado limita la libertad de expresión, incluyendo las formas concretas de expresión definidas en el artículo 20 2) del Pacto, todavía debe “justificar las prohibiciones y poner sus disposiciones en estricta conformidad con el artículo 19”<sup>16</sup>. En 2013, un grupo de expertos de alto nivel en derechos humanos, reunido bajo los auspicios del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, adoptó una interpretación del artículo 20 2)<sup>17</sup>. En el Plan de Acción de Rabat sobre la prohibición de la apología del odio nacional, racial o religioso que constituye incitación a la discriminación, la hostilidad o la violencia, los términos principales se definen como sigue:

El “odio” y la “hostilidad” se refieren a emociones intensas e irracionales de oprobio, enemistad y detestación hacia el grupo destinatario; el término “apología” debe entenderse como la intención de promover públicamente el odio hacia el grupo destinatario; y el término “incitación” se refiere a declaraciones sobre grupos nacionales, raciales o religiosos que crean un riesgo inminente de discriminación, hostilidad o violencia contra personas pertenecientes a esos grupos” (A/HRC/22/17/Add.4, apéndice, nota de pie de página 5)<sup>18</sup>.

14. En el Plan de Acción de Rabat se señalaron un total de seis factores para determinar la gravedad necesaria para tipificar como delito la incitación (*ibid.*, párr. 29):

a) El “contexto social y político que prevalecía en el momento en que se formuló y difundió el discurso”;

b) La condición del orador, “específicamente la posición del individuo u organización en el contexto de la audiencia a la que va dirigido el discurso”;

c) La intención, de manera que la “negligencia y la imprudencia no son suficientes para cometer una ofensa en virtud del artículo 20 del Pacto”, que dispone que la mera distribución o circulación no equivale a apología o incitación.

d) El contenido y la forma del discurso, en particular “el grado en que el discurso fue provocativo y directo, así como la forma, el estilo y la naturaleza de los argumentos utilizados”;

e) La extensión o el alcance del acto del discurso, como la “magnitud y el tamaño de su audiencia”, incluyendo si se trata de “un solo folleto o emisión en los principales medios de comunicación o a través de Internet, la frecuencia, la cantidad y el alcance de las comunicaciones, si la audiencia tenía medios para actuar sobre la incitación”;

f) La probabilidad, incluida la inminencia, que significa que se “pueda identificar algún grado de riesgo de daño”, en particular mediante la determinación (por parte de los tribunales, como se sugiere en el Plan de Acción) de una

<sup>15</sup> Comité de Derechos Humanos, observación general núm. 34 (2011), párr. 50.

<sup>16</sup> *Ibid.*, párr. 52, y, en el contexto del artículo 20 2) del Pacto en particular, véase párr. el 50.

<sup>17</sup> Véase, por ejemplo, Comité para la Eliminación de la Discriminación Racial, recomendación general núm. 35 (2013) sobre la lucha contra el discurso de odio racista.

<sup>18</sup> El anterior Relator Especial, Frank La Rue, definió como un factor clave en la evaluación de la incitación si había un “peligro real e inminente de violencia resultante de la expresión” (A/67/357, párr. 46). Véase también Article 19, *Prohibiting Incitement to Discrimination, Hostility or Violence* (Londres, 2012), págs. 24 y 25.

“probabilidad razonable de que el discurso logre incitar a la acción real contra el grupo destinatario”.

15. En 2013, el Comité para la Eliminación de la Discriminación Racial, el órgano de vigilancia de la Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial, siguió el ejemplo del Comité de Derechos Humanos y el Plan de Acción de Rabat. Aclaró que las palabras “debida consideración” del artículo 4 de la Convención significaban que se requiere un estricto cumplimiento de las garantías de la libertad de expresión<sup>19</sup>. En señal de convergencia de interpretaciones, el Comité subrayó que la tipificación como delito en virtud del artículo 4 debería reservarse para ciertos casos, como se indica a continuación:

La tipificación como delito de las formas de expresión racista se debe reservar para los casos más graves, que puedan probarse más allá de toda duda razonable, mientras que los casos menos graves deben tratarse por otros medios que no sean el derecho penal, teniendo en cuenta, entre otras cosas, la naturaleza y la amplitud de las repercusiones para las personas y los grupos destinatarios. La aplicación de sanciones penales debe regirse por los principios de legalidad, proporcionalidad y necesidad<sup>20</sup>.

16. El Comité para la Eliminación de la Discriminación Racial explicó que las condiciones definidas en el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos también se aplica a las restricciones en virtud del artículo 4 de la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial<sup>21</sup>. Para calificar los actos de difusión e incitación como actos punibles conforme a la ley, el Comité halló que los Estados debían tener en cuenta una serie de factores a la hora de determinar si una expresión concreta entra dentro de una categoría prohibida, como el “contenido y la forma” del discurso, el “clima económico, social y político” en cuestión en el momento de la expresión, la “posición o condición del orador”, el “alcance del discurso” y sus objetivos. El Comité recomendó que los Estados partes en la Convención consideraran “el riesgo o la probabilidad inminente de que el discurso en cuestión tenga por resultado la conducta deseada o pretendida por el emisor”<sup>22</sup>.

17. El Comité también consideró que la Convención exige la prohibición de “insultos, burlas o calumnias a personas o grupos, o la justificación del odio, el desprecio o la discriminación”, y subrayó que esas expresiones solo pueden prohibirse “cuando constituyan claramente incitación al odio o a la discriminación”<sup>23</sup>. Los términos “ridículo” y “justificación” son extremadamente amplios y, en general, están excluidos de toda restricción en virtud del derecho internacional de los derechos humanos, que protege el derecho a ofender y burlarse. Así pues, el vínculo con la incitación y con el marco establecido en virtud del artículo 19 3) del Pacto contribuye a limitar esta prohibición a la categoría más grave.

18. En el Plan de Acción de Rabat también se aclara que la tipificación como delito debe dejarse para los tipos más graves de incitación en virtud del artículo 20 2) del Pacto y que, en general, hay que considerar en primer lugar otros enfoques (A/HRC/22/17/Add.4, apéndice, párr. 34). Esos enfoques incluyen declaraciones

<sup>19</sup> Comité para la Eliminación de la Discriminación Racial, recomendación general núm. 35 (2013), párr. 19. El Comité entiende que la cláusula de debida consideración reviste especial importancia en lo que respecta a la libertad de expresión, que establece como “el principio de referencia más pertinente al sopesar la legitimidad de las restricciones de la libre expresión”.

<sup>20</sup> Comité para la Eliminación de la Discriminación Racial, recomendación general núm. 35 (2013), párr. 12.

<sup>21</sup> *Ibid.*, párrs. 4, 19 y 20.

<sup>22</sup> *Ibid.*, párrs. 15 y 16.

<sup>23</sup> *Ibid.*, párr. 13.

públicas de líderes de la sociedad que contrarresten el discurso de odio y fomenten la tolerancia y el respeto entre las comunidades, la educación y el diálogo intercultural, la ampliación del acceso a información y a ideas que contrarresten los mensajes de odio, la promoción de los principios y normas de derechos humanos y la capacitación al respecto. El reconocimiento de medidas distintas de las prohibiciones legales pone de relieve que la prohibición a menudo no será la medida menos restrictiva de que disponen los Estados que se enfrentan a problemas de discurso de odio.

### **Expresiones de odio que pueden no constituir apología o incitación**

19. Otros tipos de expresiones pueden no cumplir las definiciones o los umbrales del artículo 20 2) o del artículo 4, pero suponer, por ejemplo, apología del odio. Se plantea la pregunta de si pueden los Estados restringir la apología del odio que no constituya incitación a la discriminación, la hostilidad o la violencia. En otras palabras, la pregunta es si los Estados pueden restringir el discurso de odio cuando se define, como se hizo recientemente en la Estrategia y Plan de Acción de las Naciones Unidas sobre el Discurso de Odio, como cualquier comunicación de palabra que sea “un ataque o utilice lenguaje peyorativo o discriminatorio en relación con una persona o un grupo sobre la base de quiénes son o, en otras palabras, en razón de su religión, origen étnico, nacionalidad, raza, color, ascendencia, género u otro factor de identidad”<sup>24</sup>. Es evidente que se trata de una formulación que no llega a la altura del significado de incitación según el artículo 20 2) y el artículo 4, y aunque los Estados y las empresas deben combatir esas actitudes con educación, expresiones de condena y otras herramientas, las restricciones legales tendrán que cumplir las estrictas normas del derecho internacional de los derechos humanos.

20. Para el contenido que supone el tipo de discurso definido por la Estrategia de las Naciones Unidas sobre el Discurso de Odio, es decir, expresiones odiosas pero que no constituyen incitación, el artículo 19 3) del Pacto proporciona una orientación apropiada. Sus condiciones deben aplicarse estrictamente, de modo que cualquier restricción —y cualquier acción emprendida contra la expresión— cumpla las condiciones de legalidad, necesidad y proporcionalidad, y legitimidad. Habida cuenta de su vaguedad, una formulación como la utilizada en la Estrategia, si se pretende que guíe las prohibiciones en virtud del derecho, sería problemática por motivos de legalidad, dada su vaguedad, aunque podría servir de base para la acción política y social para contrarrestar la discriminación y el odio. Todo Estado que adopte esa definición también tendría que situar una restricción entre los motivos de limitación legítimos. En la mayoría de los casos, los derechos de los demás, según se definen en el artículo 19 3), pueden proporcionar la base adecuada, centrada en los derechos relacionados con la discriminación o la injerencia en la vida privada, o la protección del orden público. Sin embargo, en cada caso, seguiría siendo esencial que el Estado demuestre la necesidad y proporcionalidad de la acción, y cuanto más dura sea la sanción, mayor será la necesidad de demostrar una necesidad estricta<sup>25</sup>.

<sup>24</sup> En el Plan de Acción de Rabat se hace referencia a una expresión que está por debajo de los umbrales establecidos por el artículo 20 2) del Pacto, pero que “puede justificar una acción civil o sanciones administrativas” o, no da lugar a ninguna sanción, pero “sigue suscitando preocupación en términos de tolerancia, civismo y respeto de los derechos de los demás” (A/HRC/22/17/Add.4, párr. 20).

<sup>25</sup> La excepción de la moral pública prevista en el artículo 19 3) del Pacto sería improbable, pero cabe señalar que el Comité de Derechos Humanos ha aclarado que “el propósito de proteger la moral debe basarse en principios que no se deriven exclusivamente de una sola tradición” (Comité de Derechos Humanos, observación general núm. 34 (2011), párr. 32, citando al Comité de Derechos Humanos, observación general núm. 22 (1993) sobre el derecho a la libertad de pensamiento, conciencia y religión, párr. 8).

21. Algunas restricciones son específicamente desfavorables en virtud de las normas internacionales de derechos humanos. Como primer ejemplo, el Comité de Derechos Humanos observó que “la prohibición de las demostraciones de falta de respeto por una religión u otro sistema de creencias, incluidas las leyes sobre la blasfemia, es incompatible con el Pacto”, excepto en casos en que la blasfemia pueda definirse también como apología de odio religioso que constituya incitación de uno de los tipos requeridos<sup>26</sup>. Para ser claros, las leyes contra la blasfemia no cumplen con la condición de legitimidad del artículo 19 3) del Pacto, dado que el artículo 19 protege a las personas y sus derechos a la libertad de expresión y opinión; ni el artículo 19 3) ni el artículo 18 del Pacto protegen las ideas o creencias del ridículo, abuso, crítica u otros “ataques” considerados ofensivos. Varios mecanismos de derechos humanos han respaldado el llamamiento para que se revoquen las leyes relativas a la blasfemia, debido al riesgo que representan para el debate sobre las ideas religiosas y el papel que dichas leyes juegan al permitir que los gobiernos muestren preferencia por las ideas de una religión sobre otras religiones, creencias o sistemas no basados en las creencias (véase, en particular, [A/HRC/31/18](#), párrs. 59 a 61).

22. En segundo lugar, las leyes que “penalizan la expresión de opiniones sobre hechos históricos son incompatibles” con el artículo 19 del Pacto, lo que pone en tela de juicio la tipificación como delito de la negación del Holocausto y de otras atrocidades y otras leyes similares, que a menudo se justifican haciendo referencia al discurso de odio. El Comité de Derechos Humanos observó que la expresión de opiniones “erróneas” o “interpretaciones incorrectas de acontecimientos pasados” puede no estar sujeta a una prohibición general, y que las restricciones a la expresión de esas opiniones “no deberían exceder de lo autorizado” en el artículo 19 3) o “de lo prescrito en el artículo 20” del Pacto<sup>27</sup>. A la luz de estas y otras interpretaciones, la negación de la exactitud histórica de atrocidades no debe estar sujeta a sanciones penales ni a otras restricciones sin una evaluación más a fondo en virtud de las definiciones y el contexto arriba mencionados. La aplicación de cualquier restricción de ese tipo en virtud del derecho internacional de los derechos humanos debería entrañar la evaluación de los seis factores señalados en el Plan de Acción de Rabat.

23. Un tercer tipo de discurso que no suponga incitación puede consistir en una situación en la que un orador se esté dirigiendo individualmente contra una víctima identificable, pero no pretenda incitar a otros a adoptar medidas contra las personas sobre la base de una característica protegida<sup>28</sup>. Una vez más, por remisión al artículo 19 3) del Pacto, pueden restringirse esas expresiones con el fin de proteger los derechos de los demás o el orden público. Los Estados suelen restringir esas expresiones bajo la rúbrica general de “delitos de odio”, con la que las penas por una agresión física contra una persona o bienes se agrava por la motivación de odio subyacente.

24. En cuarto lugar, es importante destacar que una expresión que pueda resultar ofensiva o caracterizada por prejuicios y pueda plantear serias preocupaciones de intolerancia, a menudo no se ajusta a los criterios de gravedad que justifiquen ningún tipo de restricción. Hay toda una gama de expresiones de odio, por desagradables que sean, que no suponen la instigación o la amenaza directa, como declaraciones de prejuicios contra grupos protegidos. Esos sentimientos no estarían sujetos a la prohibición en virtud del Pacto Internacional de Derechos Civiles y Políticos o de la

<sup>26</sup> Comité de Derechos Humanos, observación general núm. 34 (2011), párr. 48. En este caso, la blasfemia sería irrelevante, solo sería pertinente la apología que constituya incitación.

<sup>27</sup> Comité de Derechos Humanos, observación general núm. 34 (2011), párr. 49. Véase Sarah Cleveland, *Hate Speech at Home and Abroad*, en Lee C. Bollinger y Geoffrey R. Stone, eds., *The Free Speech Century* (Nueva York, Oxford University Press, 2019). Véase también [A/67/357](#), párr. 55.

<sup>28</sup> Article 19, “*Hate Speech*” Explained, pág. 22.

Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial, y otras restricciones o acciones adversas requerirían un análisis de las condiciones previstas en el artículo 19 3) del Pacto. Los seis factores señalados por el Plan de Acción de Rabat para tipificar la incitación como delito también proporcionarían un valioso marco para examinar la manera de evaluar las reacciones de las autoridades públicas a esos discursos. De hecho, la ausencia de restricción no significa falta de acción; los Estados pueden (y deberían, de conformidad con la resolución 16/18 del Consejo de Derechos Humanos) adoptar medidas firmes, como condena gubernamental de los prejuicios, educación, capacitación, anuncios de interés público y proyectos comunitarios, para contrarrestar esa intolerancia y velar por que las autoridades públicas protejan a las personas contra la discriminación arraigada en este tipo de afirmaciones de odio.

25. Por último, la Convención para la Prevención y la Sanción del Delito de Genocidio exige a los Estados que tipifiquen como delito la incitación al genocidio. En algunas situaciones, como la de Myanmar, la inacción del Estado contra la incitación al genocidio puede contribuir a consecuencias muy graves para las comunidades vulnerables. Esa inacción en sí misma es condenable, y al igual que la propia incitación, debe ser combatida y castigada<sup>29</sup>.

### Normas de derechos humanos a nivel regional

26. Los sistemas de derechos humanos de Europa, América y África también articulan normas relacionadas con el discurso de odio. El Tribunal Europeo de Derechos Humanos ha subrayado que la libertad de expresión protege tipos de expresión que pueden “ofender, conmocionar o perturbar”<sup>30</sup>. Sin embargo, el Tribunal ha adoptado actitudes relativamente deferentes hacia Estados que siguen prohibiendo la blasfemia por ley sobre la base de la prohibición del discurso de odio o siguen tipificando como delito la negación del genocidio, en contraste con las tendencias observadas a nivel mundial<sup>31</sup>. A menudo el Tribunal evita por completo la cuestión del discurso de odio, basándose no en la libertad de expresión, sino en el “abuso de derechos” para declarar inadmisibles denuncias de violación<sup>32</sup>. Las normas europeas pueden variar cuando se trata de responsabilizar a los intermediarios por el discurso de odio en sus plataformas<sup>33</sup>. En cambio, las normas de la Comisión Interamericana de Derechos Humanos han tendido a ser similares a las normas internacionales antes mencionadas, mientras que las normas del sistema africano se encuentran en una etapa

<sup>29</sup> Véase, en particular, [A/HRC/39/64](#), párr. 73. En el artículo III c) de la Convención para la Prevención y la Sanción del Delito de Genocidio se pide que se tipifique como delito la “instigación directa y pública a cometer genocidio”.

<sup>30</sup> Tribunal Europeo de Derechos Humanos, *Handyside v. the United Kingdom*, demanda núm. 5493/72, fallo de 7 de diciembre de 1976, párr. 49. Véase Sejal Parmer, “The legal framework for addressing ‘hate speech’ in Europe”, presentación en la conferencia internacional para abordar el discurso de odio en los medios de comunicación, Zagreb, noviembre de 2018.

<sup>31</sup> Véase Consejo de Europa, “Discurso de odio”, ficha informativa, octubre de 2019; y Evelyn M. Aswad, “The future of freedom of expression online”, *Duke Law and Technology Review*, vol. 17 (agosto de 2018).

<sup>32</sup> Para una reseña de la práctica, véase Consejo de Europa, “Guide on article 17 of the European Convention on Human Rights: prohibition of abuse of rights”, actualizado el 31 de agosto de 2019.

<sup>33</sup> Compárese Tribunal Europeo de Derechos Humanos, Gran Sala, *Delfi AS v. Estonia*, demanda núm. 64569/09, sentencia, 16 de junio de 2015, con Tribunal Europeo de Derechos Humanos, Sección Cuarta, *Magyar Tartalomszolgáltatók Egyesülete e Index.hu Zrt v. Hungary*, demanda núm. 22947/13, sentencia, 2 de febrero de 2016. Véase también Article 19, “Responding to ‘hate speech’: Comparative overview of six EU countries”, 2018.

relativamente temprana<sup>34</sup>. En cualquier caso, no se pueden invocar normas regionales de derechos humanos para justificar una desviación de las normas internacionales de protección de los derechos humanos.

27. El Comité de Derechos Humanos ha rechazado específicamente la doctrina del margen de apreciación del Tribunal Europeo, señalando que “es preciso que el Estado parte, en cualquier caso en particular, demuestre de forma concreta la naturaleza exacta de la amenaza a cualquiera de los propósitos enumerados en el párrafo 3, que le hizo restringir la libertad de expresión”<sup>35</sup>. El Comité no concede facultades discrecionales al Estado simplemente porque las autoridades nacionales afirmen que, en general, están en mejores condiciones de comprender su contexto local.

#### **Resumen de los instrumentos de las Naciones Unidas sobre el discurso de odio**

28. El marco internacional de derechos humanos ha evolucionado en los últimos años para racionalizar lo que, a primera vista, parecen ser normas opuestas. En resumen, la libertad de expresión es un derecho legal de valor supremo para las sociedades democráticas, interdependiente respecto de otros derechos en todo el corpus del derecho de los derechos humanos y que los apoya. Al mismo tiempo, la lucha contra la discriminación, la igualdad y la participación pública equitativa y efectiva son la base de todo el corpus del derecho de los derechos humanos. El tipo de expresión que se recoge en el artículo 20 del Pacto Internacional de Derechos Civiles y Políticos y en el artículo 4 de la Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial plantea desafíos a ambos conjuntos de normas, algo que deben reconocer todos los participantes en la vida pública. Así pues, las restricciones al derecho a la libertad de expresión deben ser excepcionales, y el Estado tiene la responsabilidad de demostrar que son consistentes con el derecho internacional, las prohibiciones previstas en el artículo 20 del Pacto y en el artículo 4 de la Convención y del artículo 4 de la Convención deben estar sujetas a condiciones estrictas y restrictivas de conformidad con el artículo 19 3) del Pacto, y los Estados deberían utilizar en general instrumentos a su disposición distintos de la tipificación como delito y la prohibición, como la educación, la contrapropaganda y la promoción del pluralismo para hacer frente a todo tipo de discursos de odio.

### **III. Reglamentación del discurso de odio en línea**

#### **A. Obligaciones del Estado y reglamentación del discurso de odio en línea**

29. El estricto cumplimiento de las normas del derecho internacional de los derechos humanos protege contra los excesos gubernamentales. Como primer principio, los Estados no deberían utilizar las empresas de Internet como instrumentos para limitar expresiones que ellos mismos no puedan limitar en virtud del derecho internacional de los derechos humanos. Lo que exigen de las empresas, ya sea a través de la reglamentación o de amenazas de reglamentación, debe estar justificado en virtud del derecho internacional y ser conforme a él. Ciertos tipos de acciones contra el contenido son claramente inconsistentes con el artículo 19 3) del Pacto Internacional de Derechos Civiles y Políticos, tales como el cierre de Internet y la tipificación como delito de la disidencia política o la crítica al gobierno en línea

<sup>34</sup> Véase Comisión Interamericana de Derechos Humanos, “Discurso de odio y la incitación a la violencia”, en Comisión Interamericana de Derechos Humanos, *Violencia contra las personas lesbianas, gays, bisexuales, trans e intersex en América* (2015).

<sup>35</sup> Comité de Derechos Humanos, observación general núm. 34 (2011), párr. 36.

(véase [A/HRC/35/22](#)). Las penas impuestas a las personas que realicen un discurso de odio ilícito no deben aumentarse por el mero hecho de hacerse en línea.

30. Es útil contemplar un Estado hipotético que esté considerando una ley que responsabilice a los intermediarios en línea por no tomar medidas específicas contra el discurso de odio. Esa ley de “responsabilidad del intermediario” suele tener por objeto restringir la expresión, ya sea de los usuarios de una plataforma concreta o de la propia plataforma, en ocasiones con el fin de cumplir la obligación establecida en el artículo 20 2) del Pacto. Toda evaluación jurídica de una propuesta de ese tipo debe tener en cuenta las condiciones acumulativas establecidas en virtud del artículo 19 3) para que sea coherente con las normas internacionales relativas a la libertad de expresión<sup>36</sup>.

### Legalidad

31. El artículo 19 3) del Pacto requiere que, al imponer la responsabilidad por alojar un discurso de odio, se definan la frase en sí y los factores que intervienen en la identificación de los casos de discurso de odio. En una propuesta para imponer la responsabilidad por no eliminar la “incitación”, se debe definir el contenido de esa incitación de conformidad con el artículo 20 2) del Pacto y el artículo 4 de la Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial, en particular mediante la definición de los términos principales del Plan de Acción de Rabat antes mencionados. Si un Estado desea reglamentar el discurso de odio por motivos distintos de los previstos en el artículo 20 del Pacto y en el artículo 4 de la Convención, debe definir el contenido que es de hecho ilícito<sup>37</sup>; la precisión y la claridad que exigen el artículo 19 3) del Pacto significan que las leyes del Estado deben restringir la discreción excesiva de los agentes gubernamentales para hacer cumplir las normas o la de agentes privados para utilizar las normas con el fin de reprimir una expresión lícita, y deben disponer que se notifique debidamente a las personas para que reglamenten sus asuntos<sup>38</sup>. Sin claridad y precisión en las definiciones, existe un riesgo importante de abuso, de que se restrinja contenido legítimo y de que no se aborden los problemas en cuestión. Los Estados que aborden el discurso de odio deberían vincular estrechamente sus definiciones con las normas del derecho internacional de los derechos humanos, como las establecidas en el artículo 20 2) del Pacto.

32. Varios Estados han adoptado, o están considerando la posibilidad de adoptar, normas que obliguen a las empresas de Internet a eliminar expresiones “manifiestamente ilícitas” en un período determinado, por lo general en un plazo de 24 horas o incluso tan breve como 1 hora, o a eliminar de otro modo el contenido ilícito en un período más largo. La más conocida de estas leyes, la Ley para el

<sup>36</sup> Para una declaración sobre los principios que deberían aplicarse en el contexto de la responsabilidad de los intermediarios, véase Electronic Frontier Foundation, “Manila principles on intermediary liability”, 2015.

<sup>37</sup> Los Estados han distinguido en gran medida el contenido terrorista y “extremista” del “discurso de odio”, pero se deben aplicar los mismos principios de legalidad también a esos temas. Véase, por ejemplo, [A/HRC/40/52](#), párr. 75 e). El término “extremismo” se utiliza a menudo como sustituto del “discurso de odio”, aunque como término que no está arraigado en la ley. El término “extremismo violento” no añade mucha claridad. Los gobiernos que utilizan el término “extremismo” de buena fe en un contexto en línea parecen centrarse en el problema de la viralidad de las “ideologías terroristas y violentas extremistas” y parecen tener como objetivo contrarrestar las narrativas “extremistas” y “prevenir el abuso de Internet” (Llamamiento de Christchurch para eliminar el contenido terrorista y el extremismo violento en línea).

<sup>38</sup> Esto no excluye la posibilidad de demandas civiles que pueda presentar un individuo contra otro por actos ilícitos civiles tradicionales que tengan lugar en línea en lugar de fuera de línea. Sin embargo, en el artículo 19 del Pacto se requiere que se defina la expresión que pueda causar un daño jurídicamente reparable.

Cumplimiento de la Legislación en las Redes de Alemania, impone a las empresas la obligación de retirar de sus plataformas expresiones que sean ilegales en virtud de una serie de disposiciones específicamente señaladas del Código Penal alemán<sup>39</sup>. Por ejemplo, el artículo 130 del Código Penal prevé, entre otras cosas, la sanción de una persona que, de manera que pueda perturbar la paz pública, incite al odio contra un grupo nacional, racial o religioso o un grupo definido por su origen étnico, contra segmentos de la población o individuos por pertenecer a uno de los grupos o segmentos de la población antes mencionados o que pida la adopción de medidas violentas o arbitrarias contra ellos<sup>40</sup>. Evidentemente, la ley no define sus términos clave (especialmente “incitar” y “odio”)<sup>41</sup>, y sin embargo, a través de la Ley para el Cumplimiento de la Legislación en las Redes, impone multas importantes a las empresas que no cumplan sus disposiciones. La ley subyacente es problemáticamente vaga. Si bien la Ley para el Cumplimiento de la Legislación en las Redes debe entenderse como un esfuerzo de buena fe para hacer frente a la preocupación generalizada por el odio en línea y sus consecuencias fuera de línea, el hecho de que no se hayan definido esos términos clave socava la afirmación de que sus requisitos son compatibles con el derecho internacional de los derechos humanos.

33. Pocos Estados han involucrado a sus tribunales en el proceso de evaluación de las plataformas de discurso de odio que sean incompatibles con la legislación local, pero deberían permitir que se impusiera responsabilidad solo de acuerdo con las órdenes de tribunales independientes y con la posibilidad de apelar a petición del intermediario o de otra parte afectada por la acción (como el usuario en cuestión)<sup>42</sup>. Los gobiernos han estado aumentando la presión sobre las empresas para que actúen como árbitros del discurso de odio. El proceso de adopción también debería estar sujeto a normas rigurosas del estado de derecho, con oportunidades adecuadas para que el público haga aportaciones y audiencias y evaluaciones de las alternativas y los efectos en los derechos humanos<sup>43</sup>.

### **Necesidad y proporcionalidad**

34. Las medidas legislativas para incentivar la eliminación del discurso de odio en línea e imponer responsabilidades a las empresas de Internet por no hacerlo deben cumplir las normas de necesidad y proporcionalidad señaladas anteriormente. En los últimos años, los Estados han obligado a las empresas a retirar casi inmediatamente los contenidos, exigiendo que desarrollen filtros que impidan que se suban contenidos considerados nocivos. Se presiona para que haya instrumentos automatizados que sirvan como una forma de censura previa a la publicación. Problemáticamente, el requisito de tener un filtro de subida permitiría bloquear el contenido sin ninguna forma de debidas garantías procesales incluso antes de que se publique, invirtiendo la presunción bien establecida de que los Estados, y no los individuos, son los que

<sup>39</sup> Alemania, Ley para Mejorar el Cumplimiento de la Legislación en las Redes Sociales (Ley para el Cumplimiento de la Legislación en las Redes) (2017), art. 1 3).

<sup>40</sup> En el proyecto de ley francés sobre el discurso de odio en línea se hacen referencias similares. Véase la comunicación FRA 6/2019 y la respuesta del Gobierno de Francia, disponible en <https://spcommreports.ohchr.org/Tmsearch/TMDocuments>.

<sup>41</sup> Véase, sin embargo, Alemania, Tribunal Federal de Justicia, sentencia de 3 de abril de 2008, caso núm. 3 StR 394/07.

<sup>42</sup> El anterior Relator Especial señaló que “cualquier restricción impuesta debe ser aplicada por un organismo que sea independiente de toda influencia política, comercial o de otro tipo que no esté justificada, de un modo que no sea ni arbitrario ni discriminatorio y con las garantías adecuadas contra los abusos” (A/67/357, párr. 42).

<sup>43</sup> Véase la comunicación AUS 5/2019 y la respuesta de la Misión Permanente de Australia a la Oficina de las Naciones Unidas y otras organizaciones internacionales en Ginebra, disponible en <https://spcommreports.ohchr.org/Tmsearch/TMDocuments>.

deben justificar las restricciones a la libertad de expresión<sup>44</sup>. Como esos filtros son notoriamente incapaces de abordar el tipo de lenguaje natural que típicamente constituye un contenido odioso, pueden causar resultados significativamente desproporcionados<sup>45</sup>. Además, hay investigaciones que sugieren que tales filtros dañan desproporcionadamente a las comunidades históricamente poco representadas<sup>46</sup>.

35. La presión para que haya filtros de subida del discurso de odio (y de otros tipos de contenido) es desaconsejable, ya que conduce a que las plataformas reglamenten y eliminen contenidos legales. Aumentan el poder de las empresas con muy poca o ninguna supervisión u oportunidad de reparación. En cambio, los Estados deberían aplicar leyes y políticas que hagan que las empresas protejan la libertad de expresión y contrarresten las formas legalmente restringidas del discurso de odio mediante una combinación de características: requisitos de transparencia que permitan la supervisión pública, la aplicación de la legislación nacional por parte de autoridades judiciales independientes y otras medidas sociales y educativas en el sentido propuesto en el Plan de Acción de Rabat y en la resolución 16/18 del Consejo de Derechos Humanos.

36. Algunos Estados han adoptado medidas para hacer frente al discurso de odio ilegal por otros medios creativos y aparentemente proporcionados. Mientras que la India ha adoptado, problemáticamente, cierres de Internet como una herramienta para tratar los problemas de contenido en algunos casos, interfiriendo desproporcionadamente con el acceso de la población a las comunicaciones<sup>47</sup>, algunos estados de la India han adoptado enfoques alternativos. Uno de ellos consistía en crear líneas directas para que las personas informaran de los contenidos de WhatsApp a las autoridades encargadas de hacer cumplir la ley, mientras que otro consistía en la creación de “laboratorios de medios sociales” para controlar el discurso de odio en línea. Si bien ese tipo de enfoques requieren un desarrollo cuidadoso para que sean acordes con las normas de derechos humanos, sugieren un tipo de enfoque creativo y fuera de lo tradicional para abordar el discurso de odio sin encomendar el papel de policía de contenidos a empresas distantes<sup>48</sup>.

37. En 2019, una comisión oficial de Francia propuso un enfoque de la reglamentación de los contenidos en línea que parece proteger la libertad de expresión al tiempo que da cabida a las medidas para hacer frente al discurso de odio. Si bien en el momento de redactar el presente informe no estaba claro el estado la labor de la comisión, propuestas harían que las autoridades judiciales se ocuparan de los

---

<sup>44</sup> Comunicación OTH 71/2018, disponible en <https://spcommreports.ohchr.org/Tmsearch/TMDocuments>. Véase también la recomendación (UE) 2018/334 de la Comisión Europea, de 1 de marzo de 2018, sobre medidas para combatir eficazmente los contenidos ilícitos en línea, párr. 36, en la que se pide que se adopten medidas proactivas, recurriendo a medios automatizados, con el fin de detectar, identificar y retirar o bloquear el acceso a contenidos terroristas.

<sup>45</sup> Véase Center for Democracy and Technology, “Mixed messages? The limits of automated social media content analysis”, 28 de noviembre de 2017.

<sup>46</sup> En cuanto a las graves preocupaciones sobre la libertad de expresión planteadas en relación con los filtros de subida, véase Daphne Keller, “Dolphins in the Net: Internet content filters and the Advocate General’s *Glawischnig-Pieczek v. Facebook Ireland* opinion”, Stanford Center for Internet and Society, 4 de septiembre de 2019.

<sup>47</sup> Véanse las comunicaciones IND 7/2017 e IND 5/2016, disponibles en <https://spcommreports.ohchr.org/Tmsearch/TMDocuments>; véase también Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, “United Nations rights experts urge India to end communications shutdown in Kashmir”, comunicado de prensa, 22 de agosto de 2019.

<sup>48</sup> Chinmayi Arun y Nakul Nayak, “Preliminary findings on online hate speech and the law in India”, 8 de diciembre de 2016, pág. 11.

problemas del discurso de odio y las iniciativas de múltiples partes interesadas supervisarían las políticas de las empresas. La comisión llegó a la conclusión de que:

Por tanto, parece legítima la intervención pública para obligar a los principales agentes a asumir una actitud más responsable y protectora con respecto a nuestra cohesión social. Dadas las cuestiones de libertades civiles en juego, esa intervención debería estar sujeta a precauciones particulares. Debe 1) respetar la amplia gama de modelos de redes sociales, que son particularmente diversos, 2) imponer un principio de transparencia e inclusión sistemática de la sociedad civil, 3) aspirar a un nivel mínimo de intervención de acuerdo con los principios de necesidad y proporcionalidad y 4) recurrir a los tribunales para caracterizar la legalidad de los contenidos individuales<sup>49</sup>.

38. Este enfoque merece un mayor desarrollo y consideración, ya que aborda las cuestiones de la libertad de expresión y la cohesión social de una manera que parece permitir el respeto del derecho internacional de los derechos humanos.

### **Legitimidad**

39. La reglamentación gubernamental de los intermediarios en línea debe estar sujeta a las mismas directrices de legitimidad que las contenidas en el derecho de los derechos humanos que se aplican a todas las restricciones gubernamentales a la libertad de expresión. Como ya se ha señalado, ciertos tipos de expresiones que los Estados pueden calificar de “discurso de odio” no deben estar sujetos a una prohibición de conformidad con los artículos 19 o 20 2) del Pacto. Además, los términos jurídicos que restringen la incitación, por ejemplo, al “odio contra el régimen” o a la “subversión del poder del Estado” son motivos ilícitos de la restricción de conformidad con el artículo 19 3) del Pacto (A/67/357, párrs. 51 a 55). Las definiciones demasiado amplias del discurso de odio, por ejemplo, la prohibición de la incitación a la “discordia religiosa” o de expresiones que puedan someter a un país a actos violentos<sup>50</sup>, suelen permitir que se impongan restricciones a la expresión con fines ilegítimos o, en el caso de la reglamentación gubernamental de los intermediarios en línea, se impongan exigencias a los intermediarios que sean incompatibles con el derecho de los derechos humanos.

## **B. Moderación de los contenidos por las empresas y discurso de odio**

40. Es en las plataformas de las empresas de Internet donde el contenido odioso se propaga en línea, aparentemente impulsado por un modelo de negocio que valora la atención y la viralidad<sup>51</sup>. Las empresas más grandes despliegan “clasificadores”, utilizando software de inteligencia artificial para identificar contenidos prohibidos, con quizás un éxito solo intermitente, sobre la base de palabras y análisis específicos. Las empresas operan en diferentes jurisdicciones, y el mismo contenido en un lugar puede tener efectos diferentes en otros. El discurso de odio en línea a menudo lo hacen

<sup>49</sup> Francia, “Creating a French framework to make social media platforms more accountable: acting in France with a European vision”, informe provisional de misión presentado al Secretario de Estado de Asuntos Digitales de Francia, mayo de 2019.

<sup>50</sup> Véase la comunicación JOR 3/2018, disponible en <https://spcommreports.ohchr.org/Tmsearch/TMDocuments>.

<sup>51</sup> Véase Tim Wu, *The Attention Merchants: The Epic Scramble to Get Inside Our Heads* (Nueva York, Vintage Books, 2016).

oradores desconocidos, con amenazas coordinadas de robots, desinformación y las denominadas ultrafalsificaciones, y ataques de masas<sup>52</sup>.

41. Las empresas de Internet dan forma a las reglas de sus plataformas y la presentación pública (o marca)<sup>53</sup>. Tienen un enorme impacto en los derechos humanos, en particular, pero no solo, en lugares en que son la forma predominante de expresión pública y privada, donde una limitación de la libertad de expresión puede equivaler a un silenciamiento público o a no abordar una incitación que puede facilitar la violencia y la discriminación fuera de línea (A/HRC/42/50, párrs. 70 a 75). Las consecuencias del odio no controlado en línea pueden ser trágicas, como lo demuestra que Facebook no haya abordado la incitación contra la comunidad musulmana rohinyá en Myanmar. Las empresas no tienen las obligaciones de los gobiernos, pero sus efectos son de un tipo que les obliga a evaluar el mismo tipo de cuestiones sobre la protección del derecho de sus usuarios a la libertad de expresión<sup>54</sup>.

42. En informes anteriores se ha afirmado que todas las empresas del sector de las TIC deberían aplicar los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas e integrar los derechos humanos en sus productos por diseño y por defecto. Sin embargo, las empresas abordan el discurso de odio en sus plataformas casi por completo sin hacer referencia a las repercusiones de sus productos para los derechos humanos<sup>55</sup>. Esto es un error, ya que priva a las empresas de un marco para adoptar decisiones que respeten los derechos y articular su aplicación con los gobiernos y los particulares, al tiempo que perjudica la capacidad del público para formular reclamaciones utilizando un vocabulario mundialmente conocido. El Relator Especial reitera el llamamiento a las empresas para que apliquen políticas de derechos humanos que incluyan mecanismos para:

- a) Hacer revisiones periódicas del efecto de los productos de la empresa en los derechos humanos;
- b) Evitar efectos adversos sobre los derechos humanos y prevenir o mitigar los que surjan;
- c) Poner en marcha procesos de diligencia debida para “determinar, prevenir y mitigar su impacto en los derechos humanos” y tener un proceso para remediar el daño<sup>56</sup>.

43. Siempre habrá cuestiones difíciles sobre cómo aplicar las normas de derechos humanos de las Naciones Unidas a una amplia gama de contenidos, del mismo modo que hay cuestiones difíciles sobre las leyes nacionales y el derecho regional de los derechos humanos<sup>57</sup>. Sin embargo, la orientación mencionada anteriormente puede ayudar a configurar la protección de los derechos por las empresas en cada etapa de

<sup>52</sup> Véase Gayathry Venkiteswaran, “*Let the Mob Do the Job*”: *How Proponents of Hatred are Threatening Freedom of Expression and Religion Online in Asia* (Asociación para el Progreso de las Comunicaciones, octubre de 2017).

<sup>53</sup> Véase Kate Klonick, “The new governors: the people, rules, and processes governing online speech”, *Harvard Law Review*, vol. 131, núm. 6 (abril de 2018); y David Kaye, *Speech Police: The Global Struggle to Govern the Internet* (Nueva York, Columbia Global Reports, 2019).

<sup>54</sup> Véase A/HRC/32/38, párrs. 87 y 88; véase también Business for Social Responsibility y Foro Económico Mundial, “Responsible use of technology”, libro blanco, agosto de 2019.

<sup>55</sup> En el momento de redactar el presente informe, Facebook acababa de publicar una declaración de valores revisada en la que se indicaba que examinará las normas internacionales de derechos humanos para hacer ciertos juicios sobre las normas comunitarias. Véase Monika Bickert, “Updating the values that inform our community standards”, Facebook, 12 de septiembre de 2019.

<sup>56</sup> Principios Rectores sobre las Empresas y los Derechos Humanos: Puesta en Práctica del Marco de las Naciones Unidas para “Proteger, Respetar y Remediar” (A/HRC/17/31, anexo), principios 12 (con comentarios), 13 y 15.

<sup>57</sup> Véase Benesch, “Proposals for improved regulation”.

la moderación del contenido: desarrollo del producto, definición, identificación, acción y reparación. Las normas globales proporcionan una base firme para las empresas con usuarios globales que se comunican a través de las fronteras, y así lo exigen los Principios Rectores sobre las Empresas y los Derechos Humanos (principio 12)<sup>58</sup>.

### **Política de diligencia debida en materia de derechos humanos y examen**

44. El tratamiento del discurso de odio debe comenzar con la debida diligencia en la fase de desarrollo del producto. Lamentablemente, parece probable que pocas o ninguna de las principales empresas de Internet hayan hecho un examen de sus productos orientado a los derechos en relación con el discurso de odio, y si lo han hecho, no se ha hecho público. Sin embargo, los productos del sector de las TIC se actualizan y examinan constantemente, por lo que es fundamental que las empresas hagan evaluaciones y reevaluaciones periódicas de los efectos para determinar la forma en que sus productos infringen el disfrute de los derechos humanos. Con arreglo a los Principios Rectores sobre las Empresas y los Derechos Humanos, las empresas deben, entre otras cosas, disponer de un proceso continuo para determinar cómo el discurso de odio afecta a los derechos humanos en sus plataformas (principio 17), en particular mediante algoritmos propios de la plataforma (véase [A/73/348](#)). Deben recurrir a expertos internos e independientes en materia de derechos humanos, incluida la celebración de “consultas sustantivas con los grupos potencialmente afectados y otras partes interesadas” (principio 18). Deberían evaluar periódicamente la eficacia de sus enfoques hacia los daños a los derechos humanos (principio 20).

45. La falta de transparencia es un defecto importante en todos los procesos de moderación de contenidos de las empresas. Existe un obstáculo importante para el examen externo (académico, jurídico y de otro tipo) de las políticas relativas al discurso de odio, como se exige en virtud del principio 21: si bien las normas son públicas, los detalles de su aplicación, a nivel agregado y granular, son casi inexistentes. Por último, las empresas también deben capacitar a sus equipos de política de contenidos, a los asesores jurídicos y, en especial, a los moderadores de contenidos sobre el terreno, es decir, a los que llevan a cabo el trabajo de restricción propiamente dicho (principio 16, comentario). Como parte de la capacitación, se deben determinar cuáles son las normas del derecho de los derechos humanos que tiene por objeto proteger y promover la moderación del contenido. En particular, las empresas deben evaluar si sus normas en materia de discurso de odio infringen en la libertad de expresión, evaluando los principios de legalidad, necesidad y legitimidad señalados anteriormente.

### **Estándar de legalidad**

46. Las definiciones de las empresas sobre el discurso de odio son en general difíciles de entender, aunque las empresas varían en este sentido. En algunas empresas esas definiciones no existen y en otras son vagas. Por ejemplo, la red social rusa VK prohíbe los contenidos que propaguen o contribuyan al odio u hostilidad racial, religiosa, étnica, propaguen el fascismo o la superioridad racial o contengan materiales extremistas<sup>59</sup>. La aplicación de mensajería china WeChat prohíbe el contenido que, de hecho, o en su opinión razonable, sea odioso, acosador, abusivo, ofensivo racial o étnicamente, difamatorio, humillante para otras personas (públicamente o de otro modo), amenazante, profano o de otro modo censurable<sup>60</sup>. Otras definiciones son densas y detalladas, y hacen esfuerzos serios por explicar

<sup>58</sup> Véase Business for Social Responsibility, *Human Rights Impact Assessment: Facebook in Myanmar* (octubre de 2018).

<sup>59</sup> Véase <https://vk.com/terms>.

<sup>60</sup> Véase [www.wechat.com/en/acceptable\\_use\\_policy.html](http://www.wechat.com/en/acceptable_use_policy.html).

exactamente el tipo de contenido que constituye un discurso de odio sujeto a restricciones, pero paradójicamente, la densidad puede crear confusión y falta de claridad. Las políticas de las tres empresas estadounidenses dominantes, YouTube, Facebook y Twitter<sup>61</sup>, han evolucionado y mejorado a lo largo de muchos años, cada una de ellas ha ido superponiendo sus políticas de manera que han convergido hacia un conjunto de reglas reconociblemente similares. Sin embargo, aunque utilizan términos diferentes para señalar la restricción del contenido que “promueve” la violencia o el odio contra determinados grupos protegidos, no aclaran cómo definen la promoción, la incitación, los grupos destinatarios y otros factores. Entre otras cuestiones, es difícil encontrar en las políticas temas como la intención y los resultados (A/HRC/38/35, párr. 26).

47. Las empresas deben revisar sus políticas, o adoptar otras nuevas, teniendo en cuenta la prueba de la legalidad. Un marco conforme con los derechos humanos sobre el discurso de odio en línea se basaría en la orientación de la definición mencionada anteriormente y daría respuestas a las siguientes preguntas:

a) ¿Qué son las personas o grupos protegidos? El derecho de los derechos humanos ha determinado los grupos específicos que requieren una protección expresa. Las empresas del sector de las TIC deberían tratar de aplicar la protección más amplia posible de acuerdo con la evolución de las leyes y los entendimientos normativos. Las empresas deben dejar claro que no restringirían la promoción de un sentido positivo de identidad de grupo, en particular en el contexto de grupos históricamente desfavorecidos, reconociendo al mismo tiempo que algunas expresiones de identidad de grupo, como la supremacía blanca, pueden de hecho constituir un contenido odioso<sup>62</sup>;

b) ¿Qué tipo de discurso de odio constituye una violación de las reglas de la empresa? Las empresas deben desarrollar políticas sobre el discurso de odio teniendo en cuenta los tipos de interferencias a las que pueden enfrentarse los usuarios en la plataforma. El derecho de los derechos humanos ofrece orientación, en particular al señalar la legitimidad de las restricciones para proteger los derechos de los demás. Por ejemplo, las empresas podrían considerar cómo un discurso de odio en línea puede incitar a una violencia que amenace la vida, infrinja la libertad de expresión y el acceso a la información de otros, interfiera con la privacidad o el derecho al voto, y otros. Las empresas no están en condiciones, como los gobiernos, de evaluar las amenazas a la seguridad nacional y el orden público, y las restricciones al discurso de odio por esos motivos no deben basarse en la evaluación de las empresas sino en los ordenamientos jurídicos de los Estados, que a su vez están sujetos a las estrictas condiciones establecidas en el artículo 19 3) del Pacto;

c) ¿Existen contenidos específicos del discurso de odio que restrinjan las empresas? Las empresas deben indicar cómo prohíben, en caso afirmativo, el tipo de expresión a que se refieren el artículo 20 2) del Pacto y el artículo 4 de la Convención. Al definir esa expresión prohibida, deben basarse en los instrumentos antes mencionados. Sin embargo, la incitación es solo una parte del contenido problemático que puede constituir un discurso de odio. Las empresas deben señalar lo que incluye esa categoría además de la incitación, como ya han hecho algunas empresas a través de sus políticas en evolución. Deben hacer algo más que simplemente definir, también deberían mostrar, mediante la elaboración de una especie de jurisprudencia, exactamente el papel que juegan sus categorías en la aplicación efectiva de las normas (A/HRC/38/35, párr. 71);

<sup>61</sup> Véanse <https://support.google.com/youtube/answer/2801939?hl=en>, [www.facebook.com/communitystandards/hate\\_speech](https://www.facebook.com/communitystandards/hate_speech) and <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>.

<sup>62</sup> Artículo 19, Principios de Camden sobre la libertad de expresión y la igualdad, principio 12.

d) ¿Existen categorías de usuarios a los que no se aplican las normas sobre el discurso de odio? Las normas internacionales establecen claramente que los periodistas y otras personas que informan sobre el discurso de odio deben estar protegidos contra las restricciones de contenido o las medidas adversas tomadas contra sus cuentas. Además, la aplicación de las normas sobre el contexto del Plan de Acción de Rabat conduciría a la protección de dicho contenido. Los políticos, los funcionarios gubernamentales y militares y otras figuras públicas son otro asunto. Dada su prominencia y su posible papel de liderazgo en la incitación a comportamientos, deberían estar sujetos a las mismas normas sobre el discurso de odio que se aplican en virtud de las normas internacionales. En el contexto de las políticas sobre el discurso de odio, las personalidades públicas deben respetar las mismas normas que los demás usuarios. La evaluación del contexto puede hacer que se tome la decisión de hacer una excepción en algunos casos, cuando el contenido debe ser protegido como, por ejemplo, el discurso político. Sin embargo, es casi seguro que la incitación sea más dañina cuando es pronunciada por los líderes que por otros usuarios, y ese factor debería formar parte de la evaluación del contenido de la plataforma.

48. Cuando las reglas de las empresas difieren de las normas internacionales, las empresas deben dar una explicación razonada de antemano de la diferencia de política, de manera que se explique la variación. Por ejemplo, si una empresa decide prohibir el uso de un término despectivo para referirse a un grupo nacional, racial o religioso, que, por sí solo, no estaría sujeto a restricciones en virtud del derecho de los derechos humanos, debe aclarar su decisión de conformidad con el derecho de los derechos humanos. Además, las empresas deben estar especialmente atentas al uso indebido de sus plataformas mediante desinformación que constituya un discurso de odio; en particular en entornos de creciente tensión, las empresas deben declarar claramente sus políticas, desarrollar una comprensión global a través de la participación de la comunidad y de expertos y contrarrestar con firmeza esa incitación. Las normas internacionales de derechos humanos pueden orientar esas políticas, mientras que la viralidad de los contenidos de odio en esos contextos puede requerir una reacción rápida y una alerta temprana para proteger los derechos fundamentales.

49. Las empresas deben definir cómo determinan si un usuario ha violado las reglas relativas al discurso de odio. En la actualidad, es difícil conocer las circunstancias en las que se pueden violar las normas. Parece haber una incoherencia muy significativa en la aplicación de las normas. La opacidad de la ejecución es parte del problema. En el Plan de Acción de Rabat se señala una serie de factores aplicables a la tipificación como delito de la incitación en virtud del artículo 20 2) del Pacto, pero se deberían tener en cuenta esos factores en el contexto de las medidas de las empresas contra las expresiones. No tienen que aplicarse de la misma manera que se aplicarían en un contexto penal. Sin embargo, ofrecen un marco valioso para examinar cuándo un contenido específicamente definido —los mensajes o las palabras o imágenes que componen el mensaje— merece una restricción.

50. Las empresas pueden encontrar que hacer un análisis contextual detallado es difícil y requiere muchos recursos. Las empresas más grandes dependen en gran medida de la automatización para realizar al menos el trabajo de primera capa de identificación del discurso de odio, que requiere tener reglas que dividan el contenido en una categoría (ignorar) u otra (eliminar). Utilizan el poder de la inteligencia artificial para impulsar estos sistemas, pero los sistemas evalúan muy mal el contexto (véase [A/73/348](#)). Sin embargo, si las empresas se toman en serio la protección de los derechos humanos en sus plataformas, deben asegurarse de que definen claramente las normas y requieren una evaluación humana. Además, la evaluación humana debe ser algo más que una mera determinación de si determinadas palabras pertenecen a

una categoría determinada. Debe basarse en el aprendizaje real de las comunidades en las que puede encontrarse el discurso de odio, es decir, personas que puedan entender el “código” que a veces utiliza el lenguaje para ocultar la incitación a la violencia, evaluar la intención del orador, considerar la naturaleza del orador y de la audiencia y evaluar el entorno en el que el discurso de odio puede conducir a actos violentos. Nada de esto es posible con la inteligencia artificial por sí sola, y las definiciones y estrategias deben reflejar los matices del problema. Las empresas más grandes deberían soportar la carga de estos recursos y compartir sus conocimientos y herramientas ampliamente, como fuente abierta, para garantizar que las empresas más pequeñas y los mercados más pequeños tengan acceso a esa tecnología.

### **Necesidad y proporcionalidad**

51. Las empresas disponen de instrumentos para tratar el contenido de manera que se respeten los derechos humanos, en algunos aspectos una gama de instrumentos más amplia que la de que disponen los Estados. Esta gama de opciones les permite adaptar sus respuestas a contenidos problemáticos específicos, según su gravedad y otros factores. Pueden eliminar contenido, restringir su viralidad, etiquetar su origen, suspender al usuario pertinente, suspender a la organización que patrocina el contenido, elaborar clasificaciones para destacar el uso de contenido prohibido por parte de una persona, restringir temporalmente el contenido mientras un equipo lo examina, impedir que los usuarios monetizen su contenido, crear fricción en el intercambio de contenido, colocar advertencias y etiquetas en el contenido, dar a los individuos una mayor capacidad para bloquear a otros usuarios, minimizar la amplificación del contenido, interferir con los robots y con un comportamiento coordinado de masas en línea, adoptar restricciones geolocalizadas e incluso promover la contrapropaganda. No todas estas herramientas son apropiadas en todas las circunstancias, y pueden requerir limitaciones en sí mismas, pero muestran la gama de opciones, salvo la eliminación, de que pueden disponer las empresas en determinadas situaciones. En otras palabras, así como los Estados deben evaluar si una limitación de la libertad de expresión es el enfoque menos restrictivo, también las empresas deben hacer ese tipo de evaluación. En la evaluación, las empresas deben asumir la carga de demostrar públicamente la necesidad y proporcionalidad cuando así lo soliciten los usuarios afectados, ya sea el usuario el orador, la presunta víctima, otra persona que haya encontrado el contenido o un miembro del público.

52. Evelyn Aswad distingue tres pasos que debería dar una empresa bajo el marco de la necesidad: evaluar los instrumentos que tiene disponibles para proteger un objetivo legítimo sin interferir con el discurso mismo; determinar cuál es el instrumento que menos interfiera en el discurso; y evaluar si la medida que se selecciona logra realmente sus objetivos y demostrar que lo hace<sup>63</sup>. Ese tipo de evaluación está en consonancia con el llamamiento que se hace en los Principios Rectores sobre las Empresas y los Derechos Humanos para que las empresas se aseguren de que previenen o mitigan los daños, en particular porque ese enfoque permite a las empresas evaluar los dos tipos de daños potenciales en cuestión: las restricciones a la libertad de expresión causadas por la aplicación de sus normas y las restricciones a la libertad de expresión causadas por los usuarios que utilizan el discurso de odio contra otros usuarios o contra el público. Un enfoque que se basa en este marco permite a las empresas determinar cómo responder no solo a una incitación genuina, sino también a los tipos de expresión que son comunes en línea: el discurso que se acerca al discurso de odio y la no incitación.

---

<sup>63</sup> Aswad, “The future of freedom”, págs. 47 a 52.

## Reparación

53. Los mecanismos del derecho internacional de los derechos humanos ofrecen una gran cantidad de ideas sobre medidas correctivas del discurso de odio en línea. El artículo 2 del Pacto Internacional de Derechos Civiles y Políticos y el artículo 6 de la Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial exigen que se disponga de recursos en caso de violación de las disposiciones contenidas en ellos, y los Principios Rectores sobre las Empresas y los Derechos Humanos también requieren acceso a los recursos. En su informe de 2018 sobre la moderación del contenido, el Relator Especial destacó la responsabilidad de las empresas de reparar los efectos adversos para los derechos humanos de conformidad con los Principios Rectores ([A/HRC/38/35](#), párr. 59) y, por tanto, no es necesario que se repita en detalle en el presente informe. Resumiendo, el proceso de reparación debe comenzar con una manera eficaz de que las personas denuncien posibles violaciones de las políticas sobre el discurso de odio y debe garantizar la protección contra el uso indebido del sistema de denuncia como una forma de discurso de odio. Debe incluir un proceso transparente y accesible para apelar las decisiones de la plataforma, y las empresas deben dar una respuesta razonada que también debería ser accesible para el público.

54. Como mínimo, las empresas deben identificar públicamente los tipos de recursos que interpondrán a quienes hayan violado sus políticas sobre el discurso de odio. Puede ser que suspender al usuario sea insuficiente. Las empresas deben tener respuestas graduadas según la gravedad de la violación o la reincidencia del usuario. Deben desarrollar productos sólidos que protejan la autonomía de los usuarios, la seguridad y la libertad de expresión para reparar las violaciones. Sus enfoques pueden incluir desamplificar y desmonetizar expresiones problemáticas que no desean prohibir, por la razón que sea, pero las empresas deben, una vez más, hacer que las políticas sean claras y conocidas de antemano por todos los usuarios, sobre la base de definiciones accesibles, con advertencias para todos y la oportunidad de retirar y, de ser necesario, remediar las consecuencias de un comentario ofensivo. Pueden desarrollar programas que requieran que los usuarios suspendidos que deseen volver a la plataforma participen en tipos de reparaciones, tales como disculpas u otras formas de relación directa con otras personas a las que hayan perjudicado. Deberían contar con políticas para educar, contrarrestar, informar y capacitar. La reparación también debe incluir, para los lapsos más graves, evaluaciones de los efectos posteriores a la infracción y la elaboración de políticas para poner fin a las infracciones.

55. El Plan de Acción de Rabat y la resolución 16/18 del Consejo de Derechos Humanos también ofrecen ideas que las empresas pueden utilizar para reparar los contenidos que incitan al odio. Según el Plan de Acción de Rabat, los Estados deben velar por que las personas que hayan sufrido daños reales como resultado de la incitación al odio tengan derecho a una reparación efectiva, incluido un recurso civil o extrajudicial por daños y perjuicios. Esos recursos podrían incluir daños pecuniarios, el “derecho de corrección” y el “derecho de respuesta” ([A/HRC/22/17/Add.4](#), apéndice, párrs. 33 y 34). En su resolución 16/18, el Consejo de Derechos Humanos define instrumentos como la capacitación de funcionarios gubernamentales y la promoción del derecho de las comunidades minoritarias a manifestar sus creencias. El anterior Relator Especial instó a que se presentaran recursos de procedimiento, como “el acceso a la justicia y la eficacia de las instituciones nacionales” y sustantivos, como “reparaciones adecuadas, rápidas y proporcionales a la gravedad de la expresión, como el restablecimiento de la reputación, la prevención de la recurrencia o la compensación financiera” ([A/67/357](#), párr. 48). Sin embargo, también instó a una serie de recursos no jurídicos, que las empresas deberían evaluar y aplicar dada su responsabilidad como creadoras de

plataformas en las que prosperan contenidos odiosos. Esas medidas correctivas podrían incluir actividades educativas sobre los daños del discurso de odio y la forma en que este suele tener por objeto expulsar a las comunidades vulnerables de las plataformas (es decir, silenciarlas), promover y dar mayor visibilidad a los mecanismos de respuesta al discurso de odio, la denuncia pública del discurso de odio, como la promoción de mensajes de interés público y declaraciones de personalidades públicas, y una colaboración más estrecha con los investigadores en ciencias sociales para evaluar el alcance del problema y los instrumentos más eficaces contra la proliferación de contenidos que incitan al odio (*ibid.*, párrs. 56 a 74).

## IV. Conclusiones y recomendaciones

56. El derecho internacional de los derechos humanos debe entenderse como un marco fundamental para la protección y el respeto de los derechos humanos en la lucha contra las expresiones de odio, ofensivas, peligrosas o desfavorables. El discurso de odio en línea, la amplia categoría de expresión que se describe en el presente informe, puede dar lugar a resultados perjudiciales. Cuando se usa indebidamente la frase, puede proporcionar a los Estados malintencionados un instrumento para castigar y restringir expresiones que sean totalmente legítimas e incluso necesarias en sociedades que respetan los derechos. Sin embargo, algunos tipos de expresión pueden causar un daño real. Pueden intimidar a las comunidades vulnerables para que guarden silencio, en particular cuando se trata de apología del odio que constituye incitación a la hostilidad, la discriminación o la violencia. Si no se controla y se convierte en viral, puede crear un entorno que socava el debate público y puede perjudicar incluso a aquellos que no son usuarios de la plataforma sobre el tema. Por consiguiente, es importante que los Estados y las empresas aborden los problemas del discurso de odio con la determinación de proteger a las personas que corren el riesgo de ser silenciadas y promover un debate abierto y riguroso, incluso sobre las cuestiones más delicadas en aras del interés público.

### Recomendaciones para los Estados

57. Los enfoques de los Estados hacia el discurso de odio en línea deberían comenzar con dos premisas. En primer lugar, la protección de los derechos humanos en un contexto fuera de línea también debe aplicarse al discurso en línea. No debería haber ninguna categoría especial de discurso de odio en línea que tenga penas superiores a las del discurso de odio en otros medios. En segundo lugar, los gobiernos no deberían exigir, por medios jurídicos o amenazas extralegales, que los intermediarios tomen medidas que el derecho internacional de los derechos humanos impida a los Estados tomar directamente. En consonancia con esas bases, y con referencia a las reglas antes mencionadas, los Estados deberían hacer, como mínimo, lo siguiente en la lucha contra el discurso de odio en línea:

a) Definir estrictamente en sus leyes los términos que constituyen contenidos prohibidos en virtud del artículo 20 2) del Pacto Internacional de Derechos Civiles y Políticos y el artículo 4 de la Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial y resistir la tipificación como delito de esas expresiones, salvo en las situaciones más graves, como la apología del odio nacional, racial o religioso que constituya incitación a la discriminación, la hostilidad o la violencia, y adoptar las interpretaciones del derecho de los derechos humanos contenidas en el Plan de Acción de Rabat;

b) Examinar el derecho vigente o elaborar leyes sobre el discurso de odio para satisfacer los requisitos de legalidad, necesidad y proporcionalidad y legitimidad, y someter esa elaboración a una sólida participación pública;

c) Considerar activamente y desplegar medidas de buena gobernanza, en particular las recomendadas en la resolución 16/18 del Consejo de Derechos Humanos y el Plan de Acción de Rabat, para combatir el discurso de odio con el fin de reducir la necesidad percibida de prohibir expresiones;

d) Aprobar o revisar normas relativas a la responsabilidad del intermediario que cumplan estrictamente las normas de derechos humanos y no exigir que las empresas restrinjan expresiones que los Estados no puedan hacer directamente mediante leyes;

e) Establecer o fortalecer mecanismos judiciales independientes para garantizar que las personas puedan tener acceso a la justicia y vías de recurso cuando sufran daños reconocidos en el artículo 20 2) del Pacto o el artículo 4 de la Convención;

f) Aprobar leyes que requieran que las empresas describan en detalle y en público el modo en que definen el discurso de odio y aplican sus normas contra él, y crear bases de datos de las medidas adoptadas por las empresas contra el discurso de odio, y alentar de otra manera a las empresas a respetar las normas de derechos humanos en sus propias reglas;

g) Participar activamente en los procesos internacionales concebidos como foros de aprendizaje para hacer frente al discurso de odio.

#### Recomendaciones para las empresas

58. Las empresas han evitado durante demasiado tiempo tomar el derecho de los derechos humanos como guía para sus normas y su elaboración de normas, a pesar de las amplias repercusiones que tienen en los derechos humanos de sus usuarios y el público en general. Además de los principios adoptados en informes anteriores y en consonancia con los Principios Rectores sobre las Empresas y los Derechos Humanos, todas las empresas del sector de las TIC deberían:

a) Evaluar la forma en que sus productos y servicios afectan a los derechos humanos de sus usuarios y el público, mediante evaluaciones periódicas y públicas del impacto en los derechos humanos;

b) Adoptar políticas sobre el contenido que vinculen directamente sus normas sobre el discurso de odio con el derecho internacional de los derechos humanos en las que se indique que las normas se aplicarán de conformidad con el derecho internacional de los derechos humanos, incluidos los tratados pertinentes de las Naciones Unidas y las interpretaciones de los órganos creados en virtud de tratados y los titulares de mandatos de procedimientos especiales y otros expertos, incluido el Plan de Acción de Rabat;

c) Definir la categoría de contenido que consideren discurso de odio con explicaciones razonadas para los usuarios y el público y enfoques coherentes en todas las jurisdicciones;

d) Velar por que la aplicación de las normas relativas al discurso de odio entrañe una evaluación del contexto y del daño que imponga el contenido a los usuarios y el público, en particular garantizando que el empleo de instrumentos automáticos o de inteligencia artificial incluya a un ser humano informado;

e) Velar por que en el análisis del contenido participen las comunidades más afectadas por contenido calificado como discurso de odio y que las

**comunidades participen en la búsqueda de los instrumentos más eficaces para hacer frente a los daños causados en las plataformas;**

**f) Como parte de un esfuerzo general para hacer frente al discurso de odio, elaborar instrumentos que promuevan la autonomía individual, la seguridad y la expresión libre, e incluyan medidas para desamplificar, desmonetizar, educar, contrarrestar, informar y capacitar como alternativas, cuando proceda, al cierre de cuentas y la eliminación de contenidos.**

---