

Limitaciones y alcance de la responsabilidad de las empresas proveedoras de servicios en el discurso de odio *online*. El caso de Meta en la incitación al genocidio rohingya

Limitations and scope of the responsibility of internet service provider companies. Meta and incitement to genocide of Rohingya case

LIBIA ARENAL LORA
Profesora ayudante doctor
Universidad de Sevilla

Recibido: 29.03.2023 / Aceptado: 08.05.2023

DOI: 10.20318/cdt.2023.8053

Resumen: Este trabajo pretende analizar la relación entre el discurso de odio y la incitación al genocidio, así como el alcance de la responsabilidad legal de las empresas por estos crímenes, a partir de un caso de estudio: la contribución de la empresa Facebook (Meta) a la violencia contra la etnia *rohingya* en Myanmar. En primer lugar, se lleva a cabo una revisión de los acontecimientos sucedidos a partir de los informes elaborados por la Misión Internacional Independiente de Investigación de la ONU, desplazada al terreno desde 2017, de los que se desprenden la posible existencia de una intención genocida del Estado. En segundo lugar, explora las características y las limitaciones que tiene el discurso online para su calificación como un discurso de odio punible, no amparado por la libertad de expresión. Finalmente, explora cómo estos elementos operan para el surgimiento y la atribución de responsabilidad legal de las empresas proveedoras de servicios de internet a través del estudio los casos *Gambia c. Meta* y *Doe c. Meta* ante diferentes tribunales federales de los Estados Unidos.

Palabras clave: discurso de odio online; incitación al genocidio; responsabilidad de las empresas.

Abstract: This paper aims to analyze the relationship between online hate speech and incitement to genocide, beside the legal responsibility of companies for those crimes, based on a case study: the contribution of the company Facebook (Meta) to violence against Rohingya ethnic group in Myanmar. On the first place, a review of the events that occurred against Rohingya is carried out based on the reports prepared by the UN Independent International Investigation Mission, deployed on the ground since 2017, from which the possible existence of a genocidal intent of State was recognized. Secondly, it explores the characteristics and limitations of online speech for its classification as a hate speech crime, not protected by freedom of expression. Finally, it explores how these specific elements, characterizing online hate speech, operate for the emergence and attribution of legal responsibility of the companies that provide internet services through the study of the *Gambia v. Meta* and *Doe v. Meta* before different federal courts in United States.

Keywords: online hate speech; incitement to genocide; corporate responsibility.

Sumario: I. Introducción. II. Aproximación histórica al discurso del odio en la comisión del crimen de genocidio. III. El papel de Facebook en la difusión de mensajes de odio hacia el pueblo *rohingya*. IV. Características, limitaciones, problemas jurisdiccionales y de responsabilidad en la delimitación del discurso de odio online punible. Análisis del caso *Gambia c. Facebook*. V. De la responsabilidad civil en *Doe c. Meta* en Estados Unidos a la responsabilidad internacional de Facebook en el genocidio *rohingya*. VI. Conclusiones.

I. Introducción

1. La regulación del discurso del odio, como una conducta punible, no es una tarea sencilla ya que los límites a la divulgación de opiniones e ideas ofensivas y contrarias a valores universales, relacionados con la igualdad y la dignidad humana, encuentran su umbral en la protección a la libertad de expresión, otro valor fundamental en la construcción de sociedades democráticas y plurales.

2. A los problemas tradicionales asociados con la delimitación del discurso del odio punible se suman, en la actualidad, otros tantos que tienen su razón de ser en la dificultad de caracterizar las particularidades del discurso de odio online que se ha multiplicado de manera global a través de internet. De esta manera, hoy en día, millones de personas utilizan las redes sociales para comunicarse y publican, comparten y editan contenidos que pueden llevar mensajes de odio, incitar a la violencia y a la discriminación; y esto se puede hacer prácticamente sin control porque las empresas proveedoras de servicios, que son un actor con un enorme poder global, no tienen la obligación de vigilar todo el contenido que se genera a través de sus plataformas.

3. Desde esta perspectiva, la preocupación sobre cómo regular el discurso de odio online en el Derecho internacional, e identificar y establecer obligaciones para las empresas proveedoras de servicios, que hagan surgir su responsabilidad y evitar la impunidad, es una cuestión totalmente pertinente y de enorme relevancia.

4. Este trabajo pretende analizar, utilizando el caso de estudio de la contribución de la empresa *Facebook* al -presunto- genocidio *rohingya*, la relación entre el discurso de odio y la incitación al crimen; las características y las limitaciones que tiene el discurso de odio online para su calificación como un discurso de odio punible; y, por último, cómo en este contexto puede surgir la responsabilidad legal de las empresas proveedoras de servicios de internet, bien por la vía civil por daños, bien por la vía penal por incitación al genocidio.

5. Para abordar todas y cada una de estas cuestiones, en primer lugar, se describirá la conexión histórica entre el discurso de odio y los crímenes internacionales, en particular, se hará una mención a los acontecimientos sucedidos durante la Segunda Guerra Mundial, en el transcurso de las guerras yugoslavas entre 1991 y 2001 y en el conflicto en Ruanda en 1994. En segundo lugar, se profundizará en el caso de la violencia desatada contra la etnia *rohingya*, en el que todo apunta a que se hayan cometido crímenes de lesa humanidad y genocidio, y el papel que ha tenido la difusión del discurso de odio online, a través de la empresa *Facebook*, en el desarrollo de estos sucesos. En tercer lugar, se analizarán algunos de los problemas más relevantes para la delimitación del discurso de odio online teniendo en cuenta los estándares internacionales y cómo algunos de estos problemas se presentan y se resuelven en la práctica judicial. Para ello se realizará una presentación del caso *Gambia c. Facebook* ante un tribunal federal de los Estados Unidos. Por último, se examinará la cuestión de la responsabilidad de las empresas proveedoras de servicios por la emisión del discurso de odio online para lo que se presentará el caso *Edelson y Fields Law c. Meta Platforms* ante la Corte Suprema del Estado de California, Estados Unidos. Algunas conclusiones relevantes se expondrán para concluir este trabajo.

II. Aproximación histórica al discurso del odio en la comisión del crimen de genocidio

6. Adama Dieng, asesor de las Naciones Unidas sobre la prevención del genocidio, ha declarado con una firmeza absoluta que hay “palabras que matan”, que hay “palabras que matan más que las balas”¹.

¹ Entrevista a Adama Dieng, Asesor Especial sobre el Genocidio, sobre el discurso del odio realizada el 27 de septiembre de 2019. Disponible en: <https://www.ohchr.org/es/stories/2019/09/interview-special-advisor-genocide-adama-dieng-hate-speech>.

7. Esta circunstancia, que es extrema, ocurre cuando el discurso del odio – expresión producto de la intolerancia, prejuicios y estereotipos basados en la religión, raza, nacionalidad, origen étnico y también, ahora más presente que nunca, la identidad de género u orientación sexual²- emana de actores que tienen la capacidad de influir en las orientaciones y las acciones de amplios grupos humanos. Si bien el discurso de odio tiene unas consecuencias especialmente dramáticas cuando se encuentra instalado en el Estado, en lo político-institucional, también es especialmente grave cuando prorrumpe de la innovación y de los avances tecnológicos que controlan actores privados, como son las grandes empresas proveedoras de servicios en internet (PSI), que han facilitado que los discursos de odio online incendien las redes de forma global e instantánea.

8. La historia del discurso del odio y los actos de violencia en contra de grupos humanos por sus características particulares puede remontarse a tiempos inmemoriales –puede decirse que tiene un fundamento filosófico en el esencialismo desde el momento en que Tales de Mileto afirmó que “todo es agua”, reduciendo la riqueza y pluralidad de la realidad a un solo elemento que es considerado total y fundador³-. Sin embargo, podemos afirmar que fueron los crímenes atroces que se cometieron por parte del Gobierno nazi, durante la Segunda Guerra Mundial, los que dieron lugar a la adopción de la primera norma internacional que tipificara como crimen el discurso del odio.

9. El mismo Adam Dieng ha señalado que el Holocausto no comenzó con las cámaras de gas sino con el discurso del odio del régimen nacionalsocialista contra sus enemigos– los comunistas y los judíos-, que contribuyó a la construcción y la legitimación de todo un sistema criminal diseñado para el exterminio⁴. A partir de ahí, son universalmente conocidos los trágicos acontecimientos sucedidos durante este periodo negro de la historia.

10. Las potencias vencedoras de la Segunda Guerra Mundial -Estados Unidos, Reino Unido, la Unión Soviética y Francia- crearon el Tribunal Militar Internacional de Núremberg en 1945 para sancionar las responsabilidades de dirigentes, funcionarios y colaboradores del régimen nacionalsocialista de Adolf Hitler por los diferentes crímenes y abusos cometidos en nombre del Tercer Reich alemán⁵. Si bien el Estatuto por el que se creó el Tribunal no tipificó como un crimen de Derecho internacional los actos dirigidos a la destrucción de un grupo social –genocidio- ni tampoco la incitación pública y directa a cometer tales crímenes– el discurso del odio sí aparece recogido en el ámbito de la jurisprudencia de Núremberg. El caso contra *Julius Streicher*, fundador de un panfleto antisemita que llegó a ser un periódico de masas, sentenciado a la horca en 1946 por crímenes contra la humanidad, es un ejemplo de cómo se lograron condenas contra las personas que incendiaron el odio y alimentaron la causa general de violencia contra los judíos⁶.

11. La tipificación del crimen de genocidio y de la incitación al genocidio tuvieron que esperar hasta el 9 de diciembre del año 1948, momento en el que se adoptó en el seno de las Naciones Unidas

² La Recomendación General n. 15 sobre líneas de actuación para combatir el discurso del odio de la Comisión Europea contra el Racismo y la Intolerancia, adoptada el 8 de diciembre de 2015, ha definido como discurso del odio “el uso de una o más formas de expresión específicas —por ejemplo, la defensa, promoción o instigación del odio, la humillación o el menosprecio de una persona o grupo de personas, así como el acoso, descrédito, difusión de estereotipos negativos o estigmatización o amenaza con respecto a dicha persona o grupo de personas y la justificación de esas manifestaciones— basada en una lista no exhaustiva de características personales o estados que incluyen la raza, color, idioma, religión o creencias, nacionalidad u origen nacional o étnico, al igual que la ascendencia, edad, discapacidad, *sexo, género, identidad de género y orientación sexual*” (la cursiva es nuestra).

³ J. SOLANA DUESO, “El agua como el primer principio: Las razones de Tales de Mileto”. *CONVIVIVUM*, 2009, Núm. 22, pp. 5-23. Disponible en: <https://raco.cat/index.php/Convivium/article/view/13065>. [Consulta: 6-06-2022].

⁴ Entrevista a ADAMA DIENG, *op. cit.*

⁵ Estatuto del Tribunal Militar Internacional, Acuerdo para el establecimiento de un Tribunal Militar Internacional encargado del juicio y castigo de los principales criminales de guerra del Eje Europeo, de 8 de agosto de 1945.

⁶ La información sobre el juicio y la sentencia en contra de Julius Streicher se encuentra disponible en: <https://www.jewishvirtuallibrary.org/nuremberg-trial-judgements-julius-streicher>. Véase también sobre este tema N. ANDREWS, “Julius Streicher at Nuremberg: A Case of Justice Denied?” *British Journal of Holocaust Education*, 3:1, 1994 pp. 32-55. DOI: 10.1080/17504902.1994.11102000.

la Convención para la prevención y la sanción del crimen de genocidio⁷ (en adelante, Convención sobre el genocidio), en parte, gracias al ímprobo esfuerzo de un enorme jurista, *Rafael Lempkin*, que trabajó por darle nombre a una realidad de violencia extrema que fue el intento de exterminio del pueblo judío durante el horror del régimen nazi. La Convención definió el genocidio, un crimen de naturaleza esencialmente discriminatoria, en el artículo 2, como una serie de actos –que se enuncian en el párrafo 1⁸ - dirigidos a la destrucción total o parcial de un grupo nacional, étnico, racial o religioso; asimismo, tipificó, en el artículo 3, la incitación directa y pública a cometer este crimen⁹.

12. Ahora bien, a pesar del hecho de que la Convención se encuentre en vigor desde 1951 y de que 149 Estados sean parte de este instrumento internacional, por el que han aceptado la obligación de prevenir y sancionar el genocidio, no ha podido evitarse que otros actos de genocidio se hayan llevado a cabo en nuestra historia más reciente, en contextos en los que el discurso del odio también se ha hecho presente como germen de estos terribles acontecimientos.

13. En Yugoslavia, si bien los problemas inherentes a las relaciones interétnicas y religiosas han estado presentes en su desarrollo constitucional desde la fundación misma de la Federación, el proceso reflexivo sobre estas diferencias y la respuesta violenta para resolverlas no llegó hasta el auge de los nacionalismos y del nacionalismo extremo, en particular, del serbio. Las tesis de un programa nacional serbio –la necesidad de hacer la guerra para reparar la pérdida y la traición y crear un Estado serbio “étnicamente puro” al interior de las fronteras fijadas por “la justicia histórica”- fueron ampliamente difundidas en los medios de comunicación oficiales y no oficiales y en las revistas científicas y literarias. De esta manera, una enorme energía fue invertida en la cultura de la desesperanza del pueblo serbio, en la producción de prejuicios y de odio, en definitiva, en la demonización del otro. De hecho, puede decirse que desde el campo serbio se desarrolló sistemáticamente una campaña de odio y de difamación contra las otras etnias y religiones, que dio lugar a la limpieza étnica y al genocidio de *Sbrenica*, donde murieron 8000 personas bosnio-musulmanas¹⁰.

14. El de Ruanda ha sido uno de los mayores genocidios conocidos en la historia de la humanidad. En el año 1994, el régimen hegemónico hutu de Ruanda asesinó en 100 días entre quinientas mil y un millón de personas tutsis. Un papel especialmente importante en el genocidio lo tuvieron algunos medios de comunicación, las llamadas “emisoras del odio”. A través de la Radio Libre de las Mil Colinas y el periódico *Kangura* se incitaba casi diariamente a los miembros de la etnia hutu a exterminar a los tutsis, a los que se calificaba como de cucarachas¹¹ y de los que facilitaban información sobre dónde vivían, dónde trabajaban o dónde se escondían; matar al 70% de toda la población tutsi hubiese sido imposible únicamente mediante la acción de las milicias y las fuerzas del gobierno¹², es decir, sin la participación generalizada de la población y de sus organizaciones sociales e institucionales.

⁷ Convención sobre la Prevención y la Sanción del crimen de Genocidio, adoptada y abierta a la firma y ratificación o adhesión, por la Asamblea General en su resolución 260 A (III), del 9 de diciembre de 1948. Entrada en vigor: 12 de enero de 1951, de conformidad con el artículo XIII.

⁸ *Ídem.* Art. II.- En la presente Convención, se entiende por genocidio cualquiera de los actos mencionados a continuación, perpetrados con la intención de destruir, total o parcialmente, a un grupo nacional, étnico, racial o religioso, como tal:

- a) Matanza de miembros del grupo.
- b) Lesión grave a la integridad física o mental de los miembros del grupo.
- c) Sometimiento intencional del grupo a condiciones de existencia que hayan de acarrear su destrucción física, total o parcial.
- d) Medidas destinadas a impedir los nacimientos en el seno del grupo.
- e) Traslado por fuerza de niños del grupo a otro grupo.

⁹ *Ídem.* Art. III.- Serán castigados los actos siguientes:

- c) La instigación directa y pública a cometer genocidio.

¹⁰ M.KULLASHI, “Limpieza étnica en la ex-yugoslavia” *Praxis Filosófica*. 2011. DOI 10.25100/pfilosofica.v0i16.3045.

¹¹ D. RODRIGUEZ VÁZQUEZ, “El genocidio en Ruanda: análisis de los factores que influyeron en el conflicto”. Documento de Opinión” Instituto Español de Estudios Estratégicos (IEEE), 2017. Disponible en: https://www.ieee.es/Galerias/fichero/docs_opinion/2017/DIEEEO592017_Genocidio_Ruanda_DanielRguezVazquez.pdf.

¹² W. SCHABAS, “Hate speech in Rwanda. The road to genocide” in Lattimer, M. (Ed.). *Genocide and Human Rights* (1st ed.). Routledge, 207. DOI.org/10.4324/9781351157568.

15. Los Tribunales *ad hoc*, creados por el Consejo de Seguridad mediante las resoluciones 827, de 25 de mayo de 1993¹³, y la 955, de 8 de noviembre de 1994¹⁴, para perseguir y sancionar a los responsables de genocidio, crímenes de guerra y crímenes de lesa humanidad cometidos durante las guerras en la antigua Yugoslavia y Ruanda, han desarrollado jurisprudencia en la que se demostró la incitación pública y directa al genocidio¹⁵ y, por lo menos, se discutió sobre la incitación a la persecución dentro de los crímenes contra la humanidad¹⁶.

16. Ahora bien, nuestra historia contemporánea no solo ha sido testigo de estos trágicos acontecimientos, sino que muchas otras situaciones de violencia extraordinaria, que pueden calificarse como genocidio, han tenido lugar. De ello dan cuenta el conflicto de Darfur o la limpieza étnica en contra de los *nuer* en Sudán del Sur; la aniquilación entre milicias cristianas y musulmanas en la República Democrática del Congo; el proceso de exterminio sistemático de *yazidies*, cristianos y *chiítas* en Irak y Siria; los graves abusos cometidos sobre el pueblo *uigur* y otras minorías étnicas y religiosas en la Región Autónoma Uigur de *Sinkian* por el gobierno chino; y, finalmente, el genocidio *roghinya*, que es el caso que nos ocupa y que trataremos a continuación.

III. El papel de Facebook en la difusión de mensajes de odio hacia el pueblo rohingya

17. La historia de Myanmar desde su independencia en 1948 se ha visto empañada por décadas de conflictos armados entre el ejército, ahora denominado *Tatmadaw*, y organizaciones armadas en las regiones en las que habitan minorías étnicas. Cada conflicto ha traído muertes y daños generalizados entre la población civil: torturas y malos tratos, violencia de género, trabajo forzoso, desplazamiento y restricciones en el uso de la tierra y de los medios de subsistencia, en el acceso a la educación, a los servicios de salud y a otros servicios básicos, así como otras graves consecuencias para el pueblo. Las víctimas son predominantemente personas de minorías étnicas: *rohingya*, *kachin*, *shan*, *ta'ang*, *rakhine*, *chin*, *karen* o *kokang*¹⁷.

18. La minoría *roghinya*, de religión musulmana, se encuentra asentada en Myanmar, país de población mayoritariamente budista, prácticamente desde la primera colonia británica en el territorio, en el año 1824; pero no fue hasta 1948, tras la independencia de Myanmar del Reino Unido, que se reconoció su ciudadanía, hecho que ha generado importantes conflictos dentro del país¹⁸.

19. Los ataques generalizados y sistemáticos contra los *roghinya* se han prolongado a lo largo de 70 años. Si la Junta Militar, en el poder desde el año 1962, les negó a través de Ley Ciudadana de 1983 el estatus de ciudadano y comenzó una campaña de odio sirviéndose de normas discriminatorias, el fin de este periodo, con la incipiente transición democrática que comenzó en el año de 2010, no mejoró su suerte ya que no logró revertir unos patrones de discriminación étnica profundamente arraigados. De hecho, con la Presidencia de *Thein Sein*, a partir del 2012, es cuando las actividades más violentas del plan de limpieza étnica se recrudecieron al alinearse una serie de factores políticos -alianzas públi-

¹³ Resolución 827 (1993) Distr. GENERAL S/RES/827 (1993) 25 de mayo de 1993, aprobada por el Consejo de Seguridad en su 3217ª sesión, celebrada el 25 de mayo de 1993. UN. Doc. S/RES/827 (1993).

¹⁴ Resolución 955 (1994) Aprobada por el Consejo de Seguridad en su 3453ª sesión, celebrada el 8 de noviembre de 1994 UN. Doc. S/RES/955 (1994).

¹⁵ S. KAGAN, "The "Media case" before the Rwanda Tribunal: *The Nahimana et al.* Appeal Judgement" *Hague Justice Journal I Journal Judiciaire de la Haye* Volume/Volume 3 I Number/ N°. 1 I, 2008.

¹⁶ Sobre este último tema puede verse W. K. TIMMERMANN, "Inciting Speech in the former Yugoslavia: The *Šešelj* Trial Chamber Judgment", *Journal of International Criminal Justice*, Volume 15, Issue 1, March 2017, pp- 133–155. DOI. <https://doi.org/10.1093/jicj/mqw073>.

¹⁷ *Report of the detailed findings of the Independent International Fact-Finding Mission on Myanmar*, UN. Doc. A/HRC/42/CRP, 16 abril 2019.

¹⁸ F. J. ZAMORA CABOT/M. C. MARULLO, "El conflicto rohingya y sus proyecciones jurídicas: Aspectos destacados" *Rivista Ordine Internazionale e Diritti Umani*, n. 3, 2020.

co-privadas para la atracción de inversión extranjera- y factores económicos –acaparamiento de tierras, especulación de recursos y desarrollo de grandes obras de infraestructura-; esta circunstancia, junto con el tradicional odio religioso y racial, dio lugar a una crisis humanitaria sin precedentes para esta etnia con la huida y desplazamiento forzoso de cientos de miles de personas hacia otros países como Bangladesh, India o Tailandia¹⁹.

20. Naciones Unidas (NU), haciendo notar su preocupación por la situación en Myanmar, decidió enviar con urgencia, en 2017, a la Misión Internacional Independiente de Investigación (en adelante, la Misión), nombrada por el Presidente del Consejo de Derechos Humanos, para establecer los hechos y circunstancias de las presuntas violaciones de derechos humanos cometidas por militares y fuerzas de seguridad, en particular en el estado de *Rakhine*, con miras a garantizar la plena rendición de cuentas de los autores y la justicia para las víctimas²⁰.

21. La Misión concluyó, en el año 2019, que las prácticas generalizadas y sistemáticas de asesinatos, violaciones, arrestos arbitrarios y detención, desapariciones forzadas, tortura y otros tratos inhumanos, así como la persecución basada en la etnia y la religión podían calificarse como crímenes de lesa humanidad²¹. Asimismo, declaró que se podía inferir una intención genocida continua por parte del Estado de Myanmar y que existía un riesgo grave de que se repitiesen acciones genocidas; además, que Myanmar estaría incumpliendo su obligación de prevenir e investigar el genocidio, así como de promulgar legislación efectiva que lo tipificase como delito y lo sancionase²². Por último, la Misión declaró que el Estado contaba con la complicidad o el apoyo de otros actores estatales y no estatales con importantes intereses en la región y que las condiciones en Myanmar eran inseguras, insostenibles e imposibles para que aproximadamente un millón de desplazados regresaran a sus hogares y tierra²³.

22. Esta situación de violencia contra los *rohingya* ha sido llevada ante instancias judiciales internacionales buscando el reconocimiento de responsabilidad internacional. La Corte Internacional de Justicia (CIJ), a instancia de Gambia, ha conocido del asunto. Gambia, el 11 de noviembre de 2019, presentó una solicitud urgente para la adopción de medidas provisionales alegando que el Gobierno y el ejército de Myanmar estarían actuando en violación de la Convención sobre el genocidio²⁴. La CIJ adoptó medidas provisionales, el 23 de enero de 2020, que exigían a Myanmar prevenir el genocidio; dar pasos firmes para preservar las pruebas de los crímenes cometidos; adoptar todas las medidas posibles para evitar que militares, grupos armados irregulares y cualesquiera organizaciones y personas que pudieran estar sujetas a su control, dirección o influencia, no cometieran actos de genocidio, de incitación directa y pública a cometer genocidio, de intento de cometer genocidio o de complicidad en genocidio. Asimismo, ordenó a Myanmar presentar informes periódicos explicando qué pasos estaría siguiendo para implementar las medidas provisionales²⁵. En este momento, han concluido las audiencias públicas en las que se han presentado las objeciones preliminares presentadas por parte de Myanmar²⁶

¹⁹ Ídem., 463-467.

²⁰ Resolución A/HRC/RES/34/22, 24 de marzo de 2017.

²¹ *Detailed findings of the Independent International Fact-Finding Mission on Myanmar* UN. Doc. A/HRC/42/CRP.5. Information complementary to that contained in the official report (A/HRC/42/50), submitted to the Human Rights Council pursuant to resolution 39/2. UN. Doc. A/HRC/42/50, párr. 665.

²² Ídem, párr. 666.

²³ *Idem*, párr. 668.

²⁴ *Application instituting proceedings and request for provisional measures filed in the Registry of the Court on 11 November 2019. Application of the Convention on the Prevention and Punishment of the Crime of Genocide (The Gambia v. Myanmar)*. Disponible en: <https://www.icj-cij.org/public/files/case-related/178/178-20191111-APP-01-00-EN.pdf>.

²⁵ *Application of the Convention on the Prevention and Punishment of the Crime of Genocide (The Gambia v. Myanmar), Provisional Measures*, Order of 23 January 2020, I.C.J. Reports 2020. Disponible en: <https://www.icj-cij.org/public/files/case-related/178/178-20200123-ORD-01-00-EN.pdf>.

²⁶ *Case concerning application of the Convention of the Prevention of Genocide. The Gambia vs. Myanmar Preliminary objections of the Republic of the Union of Myanmar*. 20th January 2021. Disponible en: <https://www.icj-cij.org/public/files/case-related/178/178-20210120-WRI-01-00-EN.pdf>;

y las observaciones escritas a estas objeciones por parte de Gambia²⁷. La CIJ se pronunció sobre ellas, desestimándolas, el 22 de julio de 2022²⁸.

23. Asimismo, la Corte Penal Internacional (CPI) autorizó una investigación, en noviembre de 2019, a solicitud de la Fiscalía a la Sala de Cuestiones Preliminares III, sobre posibles crímenes de lesa humanidad de deportación forzosa de población *rohingya* de Myanmar a Bangladesh²⁹, actuación que ha sido posible porque este último Estado sí es parte del Estatuto de Roma (1998). La investigación sigue abierta hoy en día.

24. Ahora bien, y para entrar en el asunto que atañe a este trabajo, hay que señalar que el discurso del odio también ha desempeñado un papel crucial en este caso de violencia contra una minoría étnica y religiosa. De hecho, la Misión centró una parte importante de su investigación en comprobar cuál era la extensión del discurso del odio y su impacto sobre la incitación a la comisión de graves crímenes como son el genocidio y los crímenes contra la humanidad. En el informe presentado ante el Consejo de Derechos Humanos en 2018³⁰, la Misión presentó una serie de hallazgos que ponían de manifiesto la generalización del discurso del odio contra los musulmanes y, en particular, contra los *rohingya*, lo que incluía la apología prohibida del odio nacional, racial o religioso que constituye incitación a la discriminación, la hostilidad o la violencia.

25. La Misión identificó una gran cantidad de discursos de odio en todo tipo de medios - impresos, transmisiones, panfletos, CD/DVD, canciones, páginas web y cuentas de redes sociales- que construían una narrativa *antirohingya* centrada en tres temas centrales. El primero, abordaba la “amenaza musulmana” como una amenaza existencial para el país ya que identifican a los musulmanes y *rohingyas* con la inmigración ilegal masiva³¹ - “inmigrantes bengalíes ilegales”-; con terroristas -que calificaban de “yihadistas”-; y, además, con la inseguridad - “criminales” o “violadores”-³². Estas narrativas se aferraron a prejuicios y estereotipos anti-musulmanes de larga data que estaban diseñados para avivar el miedo³³. El segundo tema de narrativa anti-*rohingya*, tenía una naturaleza racial y se centraba en el peligro que esta minoría representaba para el carácter budista de la nación y la pureza racial birmana. Este tema se hacía evidente en los mensajes sobre el crecimiento de la población entre los *rohingya*, que afectaban negativamente el equilibrio demográfico -”tasas de natalidad incontrolables”, “se reproducen como conejos”, “familias extremadamente numerosas”-, y en la práctica de la poligamia y las consecuencias negativas de la Inter religión³⁴. En último lugar, el tercer tema, representaba a los musulmanes, y a los *rohingya* en particular, como una amenaza a la santidad religiosa budista. Esta narrativa se difundía a

²⁷ *Case concerning application of the Convention of the Prevention of Genocide. The Gambia vs. Myanmar. Written observations of the Gambia on Preliminary objections of the Republic of the Union of Myanmar.* 20th April 2021. Disponible en: <https://www.icj-cij.org/public/files/case-related/178/178-20210420-WRI-01-00-EN.pdf>; *Application of the Convention on the Prevention and Punishment of the Crime of Genocide (The Gambia v. Myanmar) Conclusion of the public hearings on the preliminary objections raised by Myanmar.* 28th February 2022. Disponible en: <https://www.icj-cij.org/public/files/case-related/178/178-20220228-PRE-01-00-EN.pdf>

²⁸ *Application of the Convention on the Prevention and Punishment of the Crime of Genocide (The Gambia v. Myanmar) Judgment,* 22nd July 2022. Disponible en: <https://www.icj-cij.org/public/files/case-related/178/178-20220722-JUD-01-00-EN.pdf>. La CIJ desestima las objeciones preliminares presentadas por Myanmar, en particular: entiende que Gambia es el sujeto activo del procedimiento y rechaza el abuso de derecho; declara el interés especial de Gambia en el caso en virtud de que las obligaciones que emanan de la Convención para la Prevención y la Sanción del Delito de Genocidio (1948) son *erga omnes*; afirma la irrelevancia en el proceso de la reserva al artículo VIII y declara su legitimación sobre la base del artículo IX del tratado; y, finalmente, considera sobradamente acreditada la existencia de la disputa y el conocimiento por las partes,

²⁹ *Decision Pursuant to Article 15 of the Rome Statute on the Authorization of an Investigation into the Situation in the People's Republic of Bangladesh/Republic of the Union of Myanmar,* 24th November 2019. Disponible en: https://www.icc-cpi.int/sites/default/files/CourtRecords/CR2019_06955.PDF.

³⁰ *Report of the detailed findings of the Independent International Fact-Finding Mission on Myanmar,* 17th September 2018. UN. Doc. A/HRC/39/CRP.2, párr. 130.

³¹ *Ídem,* párr. 1315.

³² *Ibidem.*

³³ *Ídem,* párr. 1314.

³⁴ *Ídem,* párr. 1316.

menudo a través de comentarios generalmente despectivos sobre las tradiciones musulmanas. Dichos mensajes giraban con frecuencia en torno a la idea de que las tradiciones y los valores musulmanes eran incompatibles u ofensivos con el budismo³⁵.

26. Una vez que la Misión verificó la presencia de este discurso generalizado en la ciudadanía, en las organizaciones de la sociedad civil, las organizaciones religiosas, las entidades políticas y sus representantes, así como dentro de los órganos e instituciones del Estado, era necesario responder a la pregunta de si este discurso se encontraba también coordinado, si era sistemático, y si existían vínculos con los brotes de violencia que se habían producido contra los *rohingya*. Sus conclusiones fueron que si bien era complicado dibujar una red completa de quiénes eran los creadores y los propagadores de los mensajes de odio, no obstante, sí pudieron identificarse una serie de grupos, en particular, grupos budistas extremistas -como 969 y *MaBaTha*-, partidos políticos e individuos nacionalistas de *Rakhine* que habían emprendido una campaña de mensajes violentos y de desinformación con la aprobación tácita del gobierno y de las fuerzas de seguridad. Esta campaña del odio se valió de publicaciones escritas, incluso de libros de textos, post en redes sociales o sermones públicos para deshumanizar y estigmatizar a los *rohingya*; y, si es difícil establecer si estas campañas orquestadas contra los musulmanes y los *rohingya* provocaron o contribuyeron a un estado general de violencia, la Misión estableció que el vínculo entre el discurso del odio y los actos de discriminación y violencia fue más que circunstancial. De hecho, se dio un patrón de retórica de incitación al odio en lugares específicos que posteriormente experimentaron violencia, y también hubo indicios de picos de incitación al odio online en torno a brotes de acciones violentas y agresiones; todos ellos patrones que, bajo recomendación de la Misión, debían ser objeto de un examen exhaustivo a futuro³⁶.

27. Estas campañas de odio, conectadas con acciones o hechos violentos, se llevaron a cabo, como ya se ha mencionado, bajo la dirección, el control, la autorización, el consentimiento, la tolerancia o la complicidad del Gobierno, de las fuerzas de seguridad y de otros actores no estatales con fuertes intereses en el Myanmar. De esta forma, las autoridades estatales no solo toleraban tales prácticas, sino que también participaban activamente en ellas y las fomentaban con la desinformación y la llamada a la acción nacionalista y patriótica³⁷. La Misión encontró declaraciones públicas profundamente problemáticas de funcionarios o representantes gubernamentales, incluidos militares, fuerzas de seguridad y políticos³⁸. Por un lado, la Misión, confirmó el alcance de este discurso en el sentido de que se había normalizado, incluso, en los debates parlamentarios sobre la grave crisis del estado de *Rakhine*, de agosto de 2017, y donde varios miembros del parlamento insistieron repetidamente en la necesidad de controlar la inmigración ilegal y prevenir la amenaza de la “Yihad” por parte de los “terroristas islámicos” en referencia a los *rohingya*³⁹. Por otro lado, un análisis sistemático de declaraciones y comunicaciones de funcionarios gubernamentales y del sector de la seguridad, así como de otros entornos oficiales, indicó que, si bien en su mayoría usaban un lenguaje menos incendiario, reflejaban y promovían las mismas narrativas defendidas por las organizaciones extremistas como *MaBaTha* y otros. La Misión, de un examen en profundidad de las comunicaciones emitidas por instituciones gubernamentales, clave en los meses posteriores a la crisis de *Rakhine*, dedujo su sustancial repercusión en el desenlace de actos de violencia al ser transmitidas por la página oficial de *Facebook* de la Oficina del Comandante en Jefe, el General en Jefe Min *Aung Hlaing*, con 2,9 millones de seguidores; la página oficial de *Facebook* del Comité de Información del Consejero de Estado, con casi 400.000 seguidores; la página oficial de *Facebook* del General en Jefe Min *Aung Hlaing*, con 1,4 millones de seguidores; y la página oficial de *Facebook* del Ministerio de Información, con 1,3 millones de seguidores⁴⁰.

³⁵ Ídem, párr. 1317.

³⁶ Ídem, párrs. 1320-1326.

³⁷ Ídem, párrs. 1340-1341.

³⁸ Ídem, párr. 1328.

³⁹ Ídem, párr. 1315.

⁴⁰ Ídem, párr. 1329.

28. Estos hechos, ayudan a traer a colación el papel de los proveedores de servicios de internet (PSI), en particular las plataformas de redes sociales, en este tipo de retórica de odio. Si bien el acceso a internet y las redes sociales se caracterizan en la era digital por democratizar y mejorar el ejercicio del derecho a la libertad de expresión, también han permitido una difusión mucho más amplia y rápida del discurso de odio. En este caso particular, las campañas y los mensajes que presentaban a los *rohingya* como violentos, deshonestos, anti-bamar, anti-budistas, inmigrantes ilegales y terroristas, en línea con las narrativas anteriormente mencionadas, encontraron una vía de difusión amplia y flexible en *Facebook*.

29. *Facebook* es, con mucho, la plataforma de redes sociales más común en uso en Myanmar⁴¹. Esta circunstancia se explica por el hecho de que solo después de la vuelta del país a la democracia, en el año 2011, las redes móviles ofrecieron acuerdos específicos que privilegiaban el acceso a esta empresa y que básicamente consistían en la instalación de servicios de internet gratuitos que tuvieran la aplicación móvil de la red social como servicio estrella⁴². Algunas investigaciones realizadas sobre el papel que ha desempeñado *Facebook* en los actos de violencia contra los *rohingya* en Myanmar han identificado miles de mensajes de odio a través de la red social⁴³ y se han logrado conectar algunas publicaciones con situaciones de tensión o de violencia concretas⁴⁴.

30. El hecho de que estos mensajes de odio y de incitación a la violencia y a la discriminación pudieran filtrarse a través de la red social *Facebook* puede deducirse básicamente de las siguientes circunstancias: el diseño del sistema y de sus algoritmos subyacentes están programados para sumar y fidelizar usuarios lo que implica la utilización de mensajes que generan mayor adhesión, entre ellos, los que incluyen violencia; ausencia de mecanismos adecuados y efectivos de moderación de contenidos y de revisores de idioma birmano; tecnología incompatible con las diferentes fuentes del idioma de Myanmar; deficiencia en los flujos de información y de respuesta ante las reclamaciones presentadas por usuarios denunciando estos hechos; problemas de automatización, incluida la exploración del uso de inteligencia artificial para identificar de manera proactiva el contenido para su revisión; y, por último, problemas también de control y de la detección de desinformación⁴⁵. En conclusión, la promoción de los mensajes de odio tuvo lugar en esta red social porque el sistema base de *Facebook*, al menos en

⁴¹ Se ha estimado que el número de usuarios activos de Facebook en Myanmar es aproximadamente de 20 millones. Sin embargo, Facebook no es la única plataforma de redes que contiene discursos de odio contra los musulmanes *rohingya*. En Twitter también han proliferado tras los ataques de agosto de 2017. Entre finales de ese mes y el 5 de septiembre, los tuits con la palabra "*rohingya*" aumentaron un 499%. A fecha de mayo de 2021, una búsqueda rápida por las redes sociales apuntadas, con palabras como "*rohingya*, propaganda, *jihad* o *hate*", siguen mostrando numerosos comentarios de ataque o defensa en referencia a esta temática. El mundo virtual se ha convertido, en este caso, en un escaparate más de la crisis y problemática que rodea a esta minoría. A.C. MORENO CANTANO/ J. M. CALVILLO CISNEROS, "Propaganda de odio contra los rohingya. Estudio de caso de Facebook y Twitter" EN B. SÁNCHEZ GUTIÉRREZ / A. PINEDA (coord.) *Comunicación política en el mundo digital: tendencias actuales en propaganda, ideología y sociedad*. 2021, ISBN 978-84-1377-562-3, pp. 792-814.

⁴² Así, como ejemplo, *Facebook* y *Myanmar Post and Telecommunications* lanzaron conjuntamente "Free Basics" and "Facebook Flex" in Myanmar. Facebook Flex es un producto que permite a los suscriptores tener una versión de texto de Facebook sin incurrir en cargos por datos. Véase en *Report of the detailed findings (...)*, *op. cit.*, párr. 1344.

⁴³ Solo la asociación *Reuters* ha identificado miles de ejemplos de posts, comentarios y fotografías de naturaleza pornográfica que atacaban a los *rohingya* en Facebook: Disponible en: <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/>

⁴⁴ Las denuncias de que las publicaciones y los mensajes a través de la red social aumentaron la discriminación y la violencia en Myanmar abundaron. De esta forma, las publicaciones de principios de 2012 sobre la presunta violación y asesinato de una mujer budista por parte de hombres *rohingya* se difundieron ampliamente y se considera que contribuyeron a la tensión y la violencia en el estado de *Rakhine* ese mismo año. De manera similar, las publicaciones sobre una presunta violación de una mujer budista por musulmanes en *Mandalay* parece que condujo a los disturbios en esta ciudad en 2014. Otro ejemplo fue la instigación de disturbios comunales, a principios de septiembre de 2017, en medio de la crisis en el estado de *Rakhine*, a través de la distribución paralela de mensajes en cadena similares, pero contradictorios, dirigidos a las comunidades musulmana y budista en *Facebook Messenger*. Cada mensaje en cadena señalaba que el otro grupo se estaba preparando para una gran violencia y animaba al destinatario a prepararse para resistir. Los grupos locales de la sociedad civil informaron de que los mensajes tuvieron un alcance nacional, causaron miedo generalizado y al menos tres incidentes violentos. Véase en A/HRC/39/CRP.2, párrs. 1347-1350.

⁴⁵ Ídem, párr. 1350.

aparición, mostró serias deficiencias para controlar los contenidos que los usuarios difundían en la red y falló en prevenir o combatir la incitación a la violencia.

31. Para hacer frente a las denuncias sobre el uso de la red social para incitar al odio en Myanmar, *Facebook*, en 2015, lanzó una versión ilustrada en idioma birmano de sus normas comunitarias, que cada usuario debía aceptar, que incluían secciones sobre el discurso de odio y la violencia y que establecían el enfoque de la empresa con respecto a las cuentas y a las noticias falsas. De acuerdo con estos estándares, *Facebook* eliminaría el contenido que violase sus políticas. Asimismo, los usuarios podían reportar dicho contenido para que sus moderadores lo revisaran⁴⁶. De esta manera, en este mismo año, *Facebook* señaló a *MaBaTha* y a los monjes *Wirathu*, *Thuseitta* y *Parmakkha* como figuras y organizaciones de odio y no se les permitió tener presencia en la red, además de que cualquier otro perfil, página, grupo o contenido que los elogiara o apoyara sería prohibido⁴⁷. El 27 de agosto de 2018, *Facebook* anunció además que había eliminado cuentas de *Instagram* y páginas de *Facebook*, seguidas por casi 12 millones de personas. Prohibió a 20 personas y organizaciones el acceso a *Facebook* en Myanmar, incluido el General en jefe *Min Aung Hlaing*. Afirmó además que había eliminado 46 páginas y 12 cuentas falsas que utilizaban páginas de opinión y noticias, aparentemente independientes, para difundir de forma encubierta mensajes de odio del ejército de Myanmar. El 5 de febrero de 2019, *Facebook* prohibió a *Aracan Army* hacer uso de su plataforma de redes sociales junto con otras tres organizaciones armadas étnicas que no habían firmado el Acuerdo Nacional de Alto el Fuego de 2012; entre estas se encontraban el Ejército de la Alianza Democrática Nacional de Myanmar, el Ejército de Independencia de *Kachin* y el Ejército de Liberación Nacional de *Ta'ang*. Además, *Facebook* declaró que había conservado todos los datos y el contenido de estas páginas con el objeto de que pudieran estar disponibles para las autoridades judiciales nacionales e internacionales debidamente constituidas a efectos de rendición de cuentas, incluido el enjuiciamiento penal de los autores de graves violaciones de derechos humanos en Myanmar⁴⁸. Igualmente, con el objeto de abordar el problema del discurso del odio y de la desinformación como una herramienta del discurso del odio, *Facebook* trabajó para dar respuesta a algunos de los problemas técnicos anteriormente mencionados contratando moderadores de contenido de lengua birmana y compatibilizando sus sistemas de identificación de idioma⁴⁹. A pesar de estos avances en el control y en la supervisión de la red social en Myanmar, el propio *Zuckerberg* reconoció que la respuesta de *Facebook* no fue “ni adecuada ni rápida”⁵⁰.

32. La Misión no tuvo dudas de que la prevalencia en general del discurso de odio en Myanmar, lo que incluía el discurso del odio online en *Facebook*, contribuyó significativamente a aumentar la tensión y a crear un clima propicio para que las personas y los grupos pudieran volverse más receptivos a la incitación y a los llamados a la violencia⁵¹. Finalmente, la Misión reiteró que era necesario realizar una investigación independiente y exhaustiva que determinara hasta qué punto la difusión de mensajes de odio en *Facebook* había aumentado la discriminación y la violencia en Myanmar de modo que se pudieran extraer lecciones apropiadas y prevenir escenarios similares⁵²; asimismo, hizo un llamamiento a la red social para que realizase esfuerzos verdaderos para combatir el uso de sus plataformas para la

⁴⁶ *Ibidem*.

⁴⁷ *Idem*, párr. 1353.

⁴⁸ *Ibidem*.

⁴⁹ En particular, ha creado un clasificador de discurso del odio en birmano, una base de datos, que ha servido para que el algoritmo opere. Véase en B. PERRIGO, *Facebook Says It's Removing More Hate Speech Than Ever Before, But There's a Catch*, TIME (Nov. 27, 2019). Disponible en: <https://time.com/5739688/facebook-hate-speechlanguages/>.

⁵⁰ *Zuckerberg*, before the United States Congress: “We’ve been too slow to deal with the hate and violence in places like Myanmar [...]. The challenges we face in a country that has fast come online are very different than those in other parts of the world, and we are investing in people, technology, and programs to help address them as effectively as possible. Disponible en: <https://docs.house.gov/meetings/IF/IF00/20180411/108090/HHRG-115-IF00-Wstate-ZuckerbergM-20180411.pdf>. (Último acceso 7 de mayo de 2022).

⁵¹ A/HRC/42/CRP.5, párr. 666.

⁵² A/HRC/39/CRP.5, párr. 1354.

incitación al odio, la violencia y la discriminación⁵³ demostrando de esta forma la preocupación por la contribución de las actividades de estas grandes y poderosas empresas de comunicación y de información a la comisión de crímenes extraordinariamente graves, que son objeto de reproche internacional, como es el caso del crimen genocidio.

IV. Características, limitaciones, problemas jurisdiccionales y de responsabilidad en la delimitación del discurso de odio online punible: análisis del caso *Gambia c. Facebook*

33. Se ha mencionado ya en el epígrafe anterior que las plataformas proveedoras de servicios de internet, como son las redes sociales, sirven para conectar a las personas y facilitar la comunicación y el acceso a la información. Estas empresas pueden mejorar la democracia y el disfrute de los derechos humanos, y han sido utilizadas de muchas formas distintas para ello. Sin embargo, como en el caso que nos ocupa, también han contribuido a la difusión vertiginosa, global y con escaso control de los discursos de odio online que incitan a la violencia y a la comisión de crímenes internacionales como es el de genocidio⁵⁴.

34. Por estas razones, el discurso del odio online, que posee características particulares, requiere un análisis propio desde una doble perspectiva: por un lado, la de los problemas de su delimitación como una conducta prohibida y sancionable desde el Derecho internacional; y, de otro lado, la de las dificultades que se presentan en el campo del surgimiento de la responsabilidad legal de los diferentes actores involucrados en su creación, emisión y difusión.

35. Para comenzar con la primera de estas cuestiones, hay que señalar que el discurso de odio no se ha definido aún de una manera general, inequívoca o autorizada en el Derecho internacional⁵⁵; si además queremos aproximarnos a una definición del discurso del odio online solo existe un único instrumento internacional que contempla estas expresiones en internet, el Protocolo Adicional al Convenio sobre la Ciberdelincuencia relativo a la penalización de actos de índole racista y xenófoba cometidos por medio de sistemas informáticos, de 28 de enero de 2003⁵⁶.

⁵³ A/HRC/42/CRP.5, párr. 666.

⁵⁴ McGonagle identifica que La accesibilidad y eficacia de Internet como medio de comunicación se debe en gran medida a la facilidad, velocidad y versatilidad con la que se puede difundir la expresión en línea. Se utiliza cada vez más para difundir el discurso de odio en diferentes formas y contextos, entre ellos: difusión de propaganda, otros tipos de (des)información, teorías de conspiración y spam de odio; intercambio de información e ideas; técnicas de optimización de motores de búsqueda, como estrategias de hipervínculos diseñadas para generar mejores resultados de búsqueda; traer a usuarios involuntarios mediante la “usurpación de nombres de dominio” y el “uso de metaetiquetas engañosas”; fines organizativos tales como la coordinación de actividades, planificación de eventos, formación, campañas de contratación; fines comerciales como la recaudación de fondos, la venta de publicaciones, videos, recuerdos y parafernalia; persecución de diversos delitos contra la persona y otras conductas delictivas o invasivas, p. la selección de (potenciales) víctimas, ciberacoso, ciberacoso. Véase en T. MCGONAGLE, “The Council of Europe against online hate speech: conundrums and challenges”. Expert Paper. Institute for Information Law (IViR). Faculty of Law, University of Amsterdam, The Netherlands, 2013. Disponible en: <http://www.ivir.nl/staff/mcgonagle.html>.

⁵⁵ El Pacto Internacional de Derechos Civiles y Políticos (1966), artículo 20.2: “Toda apología del odio nacional, racial o religioso que constituya incitación a la discriminación, la hostilidad o la violencia estará prohibida por la ley”. La Convención Internacional sobre la Eliminación de toda las formas de Discriminación Racial (1965), artículo 4,a): “Declararán 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 toda asistencia a las actividades racistas, incluida su financiación”. El Convenio Europeo de Derechos Humanos (1950) no contiene ninguna provisión específica sobre el discurso del odio y se limita a reconocer límites a la libertad de expresión recogida en el artículo 10.2 y el abuso de derecho del artículo 17. Sin embargo, el Tribunal Europeo de Derechos Humanos (TEDH) ha acogido literalmente en su jurisprudencia en concepto de la Recomendación N.º R (97) del Comité de Ministros del Consejo de Europa sobre discurso de odio (1997) que establece que: El término «discurso de odio» debe ser entendido abarcando toda forma de expresión que propague, incite, promueva o justifique el odio racial, la xenofobia, el antisemitismo u otras formas de odio basadas en la intolerancia, incluyendo la intolerancia expresada por el nacionalismo agresivo y etnocentrismo, la discriminación y la hostilidad contra las minorías, los inmigrantes y las personas de origen inmigrante”. Disponible en: https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680505d5b.

⁵⁶ El Protocolo tiene como fin armonizar la legislación penal sustantiva referente a la lucha contra la propaganda racista y xenófoba, complementado así las provisiones del Convenio. Asimismo, obliga a los Estados a tipificar en su derecho interno

36. Ahora bien, a pesar de que la regulación jurídica internacional del discurso del odio sea diversa y, por tanto, tenga distinto contenido y alcance, ha quedado establecido en diferentes instrumentos jurídicos y en la jurisprudencia internacionales que su definición como una conducta prohibida, no amparada por la libertad de expresión, se articula fundamentalmente -que no siempre necesariamente⁵⁷- en torno a la incitación directa e intencional al odio, a la violencia y a la discriminación contra grupos de personas por razón de sus características concretas de raza, nacionalidad, origen étnico y religión.

37. Esta es al menos la aproximación al discurso de odio punible del Pacto Internacional de Derechos Civiles y Políticos (PIDCP) de 1966, que señala en el artículo 20.2 lo siguiente: “(...) toda apología del odio nacional, racial o religioso que constituya incitación a la discriminación, la hostilidad o la violencia estará prohibida por la ley”. Aunque el Comité de Derechos Humanos no ha proporcionado una interpretación autorizada del artículo 20, ni de la expresión “incitación”, los Principios de *Candem* sí ofrecen una orientación útil sobre el concepto de apología de odio en el marco del artículo 20, que han reconducido hacia un estándar estricto de incitación directa en el siguiente sentido: “declaraciones sobre grupos nacionales, raciales o religiosos que puedan crear un riesgo inminente de discriminación, hostilidad o violencia contra las personas que pertenecen a dichos grupos”⁵⁸.

38. Por lo tanto, si partimos de una definición del discurso del odio, como conducta punible, que se sitúa en el umbral de la incitación intencional y directa, que crea un riesgo inminente de cometer actos de violencia y discriminación, y de que estos elementos son centrales en la identificación del es-

determinadas conductas cometidas “por medio de un sistema informático” sobre personas o grupos por razón de su raza, color, ascendencia, origen nacional o étnico o religión. Entre estas conductas se encuentran las amenazas a la comisión de delitos que son graves, los insultos, la difusión o puesta a disposición del público de material negacionista o que justifique genocidios o crímenes contra la humanidad

⁵⁷ Ahora bien, también es posible establecer un umbral más bajo para la calificación del discurso de odio como una conducta punible. De esta forma, la Convención Internacional sobre la Eliminación de todas las formas de Discriminación Racial (1965) rebaja considerablemente el estándar de la incitación sancionable al señalar que serán “actos punibles la difusión de ideas basadas en la superioridad o en el odio racial” y no solo los actos de violencia o incitación a la violencia y a la discriminación. Véase Comité para la Eliminación de la Discriminación Racial, recomendación general núm. 35 (2013). El TEDH ha considerado no solo la incitación directa a la violencia o a la discriminación para señalar conductas no protegidas por la libertad de expresión. De esta manera, la Decisión de inadmisibilidad *Garudy* c. Francia, de 14 de junio de 2003, apdo.1.i) de los fundamentos de derechos en el que el Tribunal califica el negacionismo del Holocausto de “incitación al odio contra los mismos”. De otro lado, el Convenio del Consejo de Europa para la prevención del terrorismo, de 16 de mayo de 2005, penaliza la incitación indirecta al terrorismo. El TEDH consideró incitación ciertas formas de identificación con una organización terrorista y, en particular, la difusión de mensajes alabando al autor de un atentado. Ver *Yavuz y Yaylalu* c. Turquía, de 17 de diciembre de 2013, apdo. 52. La Comisión Europea del Racismo y la Intolerancia del Consejo de Europa en el informe “General Policy Recommendation NO. 15 Combating Hate Speech” adoptado el 8 de diciembre de 2015, a pesar de adherirse en principio al referente del Plan de Acción de Rabat como punto de partida para justificar una incriminación del discurso de odio, se separa de aquél cuando afirma conscientemente ir más allá relativizando la necesidad de que la incitación sea intencional y asumiendo lo que, en nuestra opinión, sería la punición de constelaciones de casos entre el dolo eventual, y la imprudencia o, si se prefiere, el temerario desprecio a la verdad. Véase en J. M. LANDA GOROSTIZA, “Delitos de odio y estándares internacionales: una visión crítica a contracorriente”, *Revista Electrónica de Ciencia Penal y Criminología Artículos*, RECPC 22-19, 2020, pp. 1-34. La Recomendación N.º. R (97) del Comité de Ministros del Consejo de Europa sobre discurso del odio (1997) y el estándar para el discurso del odio punible se aplica a declaraciones o acciones de menor entidad e impresión que no llevan *per se* una apelación a la audiencia a cometer alguna acción hostil, discriminatoria o violenta contra el grupo objetivo. Por último, en el ámbito de la Unión Europea, la Decisión Marco 2008/913/JAI del Consejo, de 28 de noviembre de 2008, relativa a la lucha contra determinadas formas y manifestaciones de racismo y xenofobia mediante el derecho penal prevé que los Estados adopten medidas necesarias para garantizar el castigo penal de la incitación pública a la violencia o al odio (art. 1.1.a), pero también de la difusión o reparto de escritos, imágenes y otros materiales (art.1.1.b) y la apología pública, la negación y la trivialización de internacionales cuando puedan incitar a la violencia o al odio. Ahora bien, la norma reconoce la capacidad de los Estados de modular la respuesta punitiva castigando solo las que cumplan con las exigencias de perturbación de orden público o que sean amenazadoras, abusivas o insultantes (art 2).

⁵⁸ Véanse Principios de *Candem*. Principio 12.1.iii, p. 10. De esta definición se desprende, por un lado, que las declaraciones hacen surgir un riesgo real que debe poder ser identificado y, por otra parte, la inminencia de dicho riesgo que debe establecerse sobre un base casuística en la que la duración del tiempo transcurrido entre la acción y los actores incitados no debe ser tanta para impedir que el emisor no sea razonablemente responsable del resultado final, es decir, que se dirija con una llamada a la acción inequívoca a un audiencia concreta, en un tiempo y lugar concretos. Sobre esta cuestión puede verse Véase también G. ROLLNERT LIERN, “El discurso del odio: Una lectura crítica internacional” *Revista Española de Derecho Constitucional*, 115, 81-109. Disponible en: <https://doi.org/10.18042/cepc/redc.115.03>.

tándar para las conductas típicas⁵⁹, la cuestión que se plantea es cómo intervienen estos elementos en la delimitación del discurso de odio online.

39. En primer lugar, y a propósito de la intención para la identificación de un delito de odio⁶⁰, hay que observar que este elemento subjetivo se exige tanto respecto de la difusión del mensaje como de la realización de los actos hacia los que se pretende mover a la audiencia o a los destinatarios de sus mensajes. Dicho de otro modo, el autor debe adherirse intencionalmente a la distribución, la circulación o la puesta a disposición de terceros de un mensaje online que incite a la violencia y a la discriminación, así como compartir el propósito de que el efecto de su difusión vaya a ser efectivamente promoverlos⁶¹. De esta manera, por un lado, la valoración de la intención del emisor en el discurso de odio online presenta singularidades debido a los rasgos de desinhibición, libertad, automatismo de la respuesta y de la sensación de anonimato que producen las especificidades del medio, en particular, el de las redes sociales, lo que conduce a que sus códigos expresivos no siempre sean claros o sean conocidos por los emisores de los mensajes. Esto significa que no siempre la difusión de un mensaje de odio implica que exista una adhesión estricta al contenido de este; tampoco que se difunda con la intención de incitar, fomentar o promover el odio y la discriminación. De esta forma, en este contexto complejo y volátil de la comunicación online, hay que evaluar el propósito del emisor caso a caso, y partir del conjunto de sus mensajes y no de estos aisladamente⁶². Asimismo, la intención exige la activación de una relación triangular entre el objeto, el sujeto del acto y la audiencia⁶³. Sin embargo, la compleja relación que se establece entre los individuos y los grupos en internet, especialmente en las redes sociales, conlleva importantes problemas de delimitación de las conductas típicas. De este modo, ha de tenerse en cuenta el carácter privado o público de las comunicaciones en relación con la intención del autor de dirigir un mensaje de odio hacia una audiencia predeterminada, el efecto en los grupos sobre los que se proyecta y el riesgo que representa.

40. En este sentido, cabe señalar que no siempre es posible diferenciar entre las manifestaciones o los mensajes de odio que se producen en el ámbito privado y público. Así, si se consideran comunicaciones o expresiones estrictamente privadas -quedando excluidas de las conductas tipificables -aquellas en las que se identifique como criterio principal la intención del emisor de que el mensaje sea recibido por un destinatario predeterminado, por otra parte, los mensajes que están dirigidos a grupos de

⁵⁹ HRC, Annual report of the United Nations High Commissioner for Human Rights and reports of the Office of the High Commissioner and the Secretary-General, 11th January 2013. UN Doc. A/HRC/22/17/Add.4 El Plan de Acción de Rabat recomienda una prueba de umbral que consta de seis parámetros para identificar el discurso de odio punible de acuerdo al Pacto que tienen en cuenta (1) el contexto social y político, (2) la categoría del hablante, (3) la intención de incitar a la audiencia contra un grupo determinado, (4) el contenido y la forma del discurso, (5) la extensión de su difusión, y (6) la probabilidad de causar daño, incluso de manera inminente. Véase La Recomendación General 35 del Comité para la eliminación de la discriminación racial establece estos mismos criterios del Plan de Acción de Rabat para identificar los discursos de odio punible. Véase CEDR, Recomendación General 35. La lucha contra el discurso racista, Aprobada por el Comité en su 83^o período de sesiones (12 a 30 de agosto de 2013). UN Doc. CERD/C/GC, párr. 14.

⁶⁰ El Protocolo Adicional al Convenio sobre la Ciberdelincuencia, de 28 de enero 2003, y la Decisión Marco 2008/913/JAI del Consejo, de 28 de noviembre de 2008, relativa a la lucha contra determinadas formas y manifestaciones de racismo y xenofobia mediante el derecho penal, establecen expresamente la responsabilidad penal por estos actos cuando se cometen intencionalmente

⁶¹ G. ROLLNERT LIERN, “Redes sociales y discurso del odio: perspectiva internacional” *Revista de Internet, Derecho y Política*, n.º.31, UOC, 2020, p. 56

⁶² Ídem.

⁶³ Temperman ha representado mediante la expresión «triángulo del odio» las interacciones existentes entre los distintos elementos concurrentes para superar el umbral de aplicación del art. 20.2 PIDCP: «Lo que mínimamente necesitamos según el art. 20.2 es un promotor [*advocator*] que exprese su intenso [...] odio hacia un grupo diana [*target group*] e incite a un tercero, la audiencia del incitador, a actos de discriminación, hostilidad o violencia *vis-à-vis* el grupo diana» (2014: 15). Según esta visión del umbral, los vértices del triángulo serían el incitador, el grupo diana y la audiencia, y la presencia de esta última determinaría la existencia de incitación. Para que el discurso pueda ser prohibido conforme al art. 20.2, los lados del triángulo deben ajustarse también a ciertas exigencias mínimas, tanto desde la perspectiva del contenido del mensaje —odio «intenso20» y llamada expresa a la acción— como del contexto —potencialmente generador de probabilidad o riesgo de hostilidad, discriminación o violencia hacia el grupo diana— en el que se produce la expresión de odio. Ver en J. TEMPERMAN, “Blasphemy versus incitement: an international law perspective.” En C. BENEKE, C. GREY y D. NASH (eds.). *Profane: Sacrilegious Expression in a Multicultural Age*, Oakland: University of California Press, pp. 401-425.

destinatarios, que por el número de integrantes y las relaciones que mantienen entre ellos ya no pueden considerarse estrictamente cerrados, sino semipúblicos o semiprivados, presentan serias dudas respecto de su calificación⁶⁴. En este marco de relaciones se produce una difícil casuística que, indudablemente, repercute en el nivel de riesgo de que se produzcan las conductas a las que se incita.

41. El riesgo equivale a la probabilidad inminente de que el discurso en cuestión tenga por resultado la conducta deseada o pretendida de odio, violencia o discriminación⁶⁵. Para valorarlo hay que tener en cuenta aspectos importantes como la relevancia que tiene el actor que emite el discurso debido a su posición, la magnitud y el tamaño de la audiencia, el contenido y la forma del discurso y el medio a través del que emite el mensaje y el alcance de sus declaraciones⁶⁶. Respecto a la evaluación del riesgo en el discurso del odio online, es inevitable tener en cuenta el impacto potencial de las publicaciones en internet a causa de la alta publicidad de los mensajes online. Ahora bien, si puede presuponerse que una mayor publicidad conlleva un mayor riesgo de que la incitación o la propagación del odio den lugar a acciones reales de violencia o discriminación⁶⁷, hay que considerar si el contexto de fugacidad e incluso de itinerancia de las redes sociales, además del de la ausencia de una audiencia concreta, son apropiados para crear un riesgo inminente. En este sentido, puede señalarse que el estándar de riesgo inminente de la doctrina de *Brandenburg* sobre incitación⁶⁸ es difícilmente aplicable a las redes sociales por requerir tanto de una inmediatez temporal entre las declaraciones y el peligro como de una audiencia concreta que ejecute las acciones que promueve. Este tipo de riesgo podría cumplirse cuando se producen llamadas inmediatas a la violencia o a la discriminación, pero no cuando sus efectos se producen a largo plazo⁶⁹. La fugacidad de las redes sociales hace pensar que solo en pocas ocasiones el mensaje de odio online podría cumplir con el criterio de riesgo inminente. Por esta razón, algunos autores, basados en precedentes judiciales, defienden la utilidad de la doctrina de las “amenazas verdaderas” para la delimitación de las conductas en el discurso online, ya que no requiere de la inminencia del riesgo, sino de la intención intimidatoria del emisor, es decir, la de provocar temor en el destinatario sin la necesidad de que se produzca el resultado violento⁷⁰.

42. Dicho lo anterior, puede afirmarse que, si la delimitación del discurso de odio ya es problemática en sí misma, la delimitación del odio online presenta dificultades añadidas debido a sus particularidades, asunto que se viene a manifestar también en materia de surgimiento de la responsabilidad legal y de la definición de la jurisdicción competente.

43. Respecto de la cuestión de la responsabilidad legal, cabe señalar que una multiplicidad de diferentes actores, al menos potencialmente, podrían estar involucrados en la creación y difusión de contenido odioso -crearlos u obtenerlos, publicarlos, desarrollarlos, alojarlos, editarlos o facilitar de otro modo su difusión, accesibilidad o recuperabilidad-. Además, a diferencia de la relación que se entabla en otros medios de comunicación tradicionales, en el discurso online no existe normalmente un acuerdo entre el autor y el responsable editorial, ya que se produce lo que se llama la auto publicación –publicaciones directas sin acuerdo del editor- hecho que limita el establecimiento de restricciones legales a las decisiones editoriales en defensa de la publicación responsable⁷¹. En este contexto de la publicación online, la atribución de responsabilidad por un mensaje de odio que incite a la violencia, en el que estén involucrados

⁶⁴ G. ROLLNERT LIERN, “Redes sociales y discurso del odio (...)” *op. cit.*, p. 56.

⁶⁵ De acuerdo con los Principios de *Candem* de 2009 el elemento decisivo en la incitación es la generación de un riesgo inminente de discriminación, hostilidad y violencia. Principio 12.1.iii. Comité para la Eliminación de la Discriminación Racial, Recomendación General núm. 35 (2013) párrs. 15 y 16.

⁶⁶ Plan de Acción de Rabat, *op. cit.*

⁶⁷ Tribunal Europeo de Derechos Humanos, caso *Savva Terentyev v. Rusia*, sentencia 28 de agosto de 2018, apartado 79.

⁶⁸ J. GRAU ÁLVAREZ, “La libertad de expresión y discurso del odio. Estudio comparado de la jurisprudencia de Estados Unidos y Europa” *ICADE. Revista de la Facultad de Derecho*, nº 111, enero-junio 2021, pp. 1-33. DOI: 10.14422/icade. i111 y2021.003

⁶⁹ G. ROLLNERT LIERN, “Redes sociales y discurso del odio...”, *op. cit.*, p. 8.

⁷⁰ Ídem.

⁷¹ C. O’REGAN, “Hate speech online: ¿an (intactable) contemporary challenge?” *Current Legal Problems*, Volume 71, Issue 1, 2018, pp. 403–429. DOI: <https://doi.org/10.1093/clp/cuy012>.

diferentes actores con distintas contribuciones, se vuelve extremadamente complejo, ya que habría que establecer distinciones legales relevantes entre los diferentes tipos de usuarios en virtud de los distintos niveles de participación, control editorial y responsabilidad en la toma de decisiones finales, cuestión que entraña enormes dificultades⁷².

44. Unido a estos problemas relativos a la atribución de responsabilidad a los emisores de mensajes de odio, el discurso online se beneficia del *forum-shopping* que implica que los emisores disfrutan de una enorme movilidad para ofrecer contenidos a través de proveedores de servicios de internet (ISP), basados en una jurisdicción de su elección con una regulación que consideren más favorable o tolerante con el discurso de odio⁷³. También es una práctica común que los sitios web de odio, que han sido bloqueados o prohibidos en una jurisdicción, se trasladen posteriormente a otra más favorable, lo que viene a indicar que el bloqueo o la prohibición no pueden ser consideradas como medidas totalmente efectivas contra dichos sitios⁷⁴. De igual manera, mientras que diferentes jurisdicciones tienen diferentes leyes que rigen el discurso de odio, diferentes ISP dentro de una jurisdicción determinada también pueden tener diferentes políticas sobre el discurso de odio. Los términos de servicio o los contratos de uso, las pautas de la comunidad y las políticas de aviso y eliminación de las ISP pueden variar en cuanto a su detalle o severidad con respecto a la incitación al odio. Además, un factor importante que hay que considerar es que los actores relevantes suelen ser actores privados y, en consecuencia, su participación en la regulación de contenido podría verse potencialmente como censura privada⁷⁵. En definitiva, el carácter eminentemente interjurisdiccional del discurso del odio online⁷⁶ constituye un enorme desafío para abordar los conflictos que puedan surgir relativos al foro competente y al derecho aplicable en materia de delimitación de las conductas y de la responsabilidad de los emisores de mensajes de odio y de las empresas proveedoras de servicios, más aún, si tenemos en cuenta que la regulación de estas conductas se desarrolla fundamentalmente en los derechos internos, y que pueden ser muy dispares en cuanto a su contenido y alcance.

45. Algunos de los problemas mencionados – intención del autor, la caracterización de los mensajes como públicos o privados a efectos legales, la divulgación de los mensajes por los proveedores o intermediarios- se han visto reflejados en la demanda que Gambia ha presentado contra *Facebook* ante el Tribunal Federal del Distrito de Columbia, Estados Unidos (EE.UU.), el 22 de septiembre de 2021⁷⁷.

46. Gambia solicitó al tribunal, de conformidad con la 28 USC §1782 sobre la asistencia a tribunales extranjeros y a litigantes ante dichos tribunales⁷⁸, por su relación y utilidad en el caso Gambia c. Myanmar ante la CIJ, que ordenara a la empresa de redes sociales *Facebook* la presentación de material relevante para el caso. En particular, Gambia solicitó que esta empresa presentara documentos relativos

⁷² Ídem, p.28.

⁷³ Este es el caso de EE.UU. donde la tradición de la libertad de expresión cultivada por la Primera Enmienda ha resultado en una presunción muy fuerte de protección constitucional para la libertad de expresión. The First Amendment to the US Constitution reads: “Congress shall make no law [...] abridging the freedom of speech, or of the press [...]”. Véase A. VALERO HEREDIA, “Los discursos del odio. Un estudio jurisprudencial” *Revista Española de Derecho Constitucional*, 110, 2017, pp. 305-333. DOI: <https://doi.org/10.18042/cepc/redc.110.11>

⁷⁴ T. MCGONAGLE, *The Council of Europe against online hate speech...*, *op. cit.*, p. 27.

⁷⁵ Ídem, pp. 27 y 28. McGonagle en su informe expone el caso de Google y su Informe de Transparencia de 2010 que adolece de lagunas importantes en el sentido que los datos que arroja permanecen en el nivel de las macroestadísticas y no ofrece información real sobre cómo se tratan las solicitudes de eliminación relacionadas con el supuesto discurso de odio en términos de revisión interna sustantiva. De hecho, en el informe se indica que el discurso de odio había representado el 0% del número total de solicitudes en virtud de órdenes judiciales y solo el 2% del total de solicitudes realizadas por el ejecutivo o la policía. Este informe se encuentra disponible en: <http://www.google.com/transparencyreport/removals/government/>

⁷⁶ C. O'REGAN, “Hate speech online...”, *op. cit.*

⁷⁷ United States District Columbia, In re Application Pursuant to 28 U.S.C. § 1782 of The Republic of the Gambia, Petitioner, v Facebook, Inc., Respondent, Civil Action No. 20-mc-36-JEB-ZMF Facebook, Inc., Respondent. Civil Action No. 20-mc-36-JEB-ZM, p. 31.

⁷⁸ 28 U.S. Code § 1782 - Assistance to foreign and international tribunals and to litigants before such tribunals. Disponible en: 28 U.S. Code § 1782 - Assistance to foreign and international tribunals and to litigants before such tribunals | U.S. Code | US Law | LII / Legal Information Institute (cornell.edu)

a cuentas y páginas que había eliminado y que, por tanto, se encontraban inaccesibles para los usuarios y el público general, pero que mantenían almacenados electrónicamente.

47. El Tribunal reconoció que Gambia era parte interesada, de acuerdo con su condición de demandante ante la CIJ, y ordenó a *Facebook* que presentará las comunicaciones públicas y privadas con material con incitación al odio o de contenido violento, así como aquellos documentos asociados con la investigación interna que la propia empresa había realizado; documentos que al mismo tiempo explicarían cómo había identificado el contenido eliminado y lo conectó con las cuentas personales en *Facebook* de los funcionarios de Myanmar, en lo que la misma empresa calificó como una “campaña coordinada de odio contra los *rohingya*”⁷⁹.

48. En cuanto a la primera de las peticiones planteadas, es necesario explicar que *Facebook* alegó que los documentos eliminados, pero retenidos electrónicamente, se encontraban sujetos a la regla de no divulgación de la 18 USC §2702 (SCA)⁸⁰, sobre divulgación voluntaria de comunicaciones o registros de clientes -comunicaciones almacenadas electrónicamente-. Sin embargo, el tribunal concluyó que este tipo de documentos (“delete-but-retained”) no se encontraban protegidos por la prohibición de no divulgación. El Tribunal se preguntó si los documentos y comunicaciones vinculadas a cuentas específicas de *Facebook*, y que ésta eliminó de la plataforma, estaban “en almacenamiento electrónico”, es decir bajo “almacenamiento temporal” o “almacenamiento de respaldo” y, por tanto, sujetos a la regla de no divulgación de la SCA. Para responder a esta cuestión, el Tribunal se centró en analizar si el contenido que un proveedor de servicios elimina, pero que retiene electrónicamente, es un “almacenamiento de respaldo”, concluyendo en este caso que no lo era. El Tribunal razonó que el “almacenamiento de respaldo” requiere que exista una versión “original” correspondiente de los documentos o información solicitada; por tanto, una copia de archivo mantenida para algún otro propósito, en ausencia del original, no cumple ninguna función de respaldo. En vista de que *Facebook* eliminó el contenido en cuestión del sitio, no podría existir lo que la SCA califica como una copia de respaldo y la regla de no divulgación no era, por tanto, aplicable⁸¹.

49. No obstante, lo que es interesante al caso es que el Tribunal pudo llegar a esta misma conclusión a partir de una de las excepciones que presentó Gambia, previstas en esta misma Ley, basada en el “consentimiento legítimo”, que implica que un proveedor de servicios “puede divulgar el contenido de una comunicación”⁸² cuando ese consentimiento, de manera explícita o implícita, se presta; cuestión que, por otro lado, se encuentra íntimamente ligada con la calificación del material como público o privado.

50. El Tribunal concluyó que un usuario que publica una comunicación, sobre una base razonable para saber que estará disponible para el público, otorga un consentimiento implícito a su divulgación⁸³. En el caso que nos ocupa, una gran parte del contenido buscado por Gambia había sido de acceso

⁷⁹ United States District Columbia, In re Application Pursuant to 28 U.S.C. § 1782 of The Republic of the Gambia, Petitioner, v Facebook, Inc., Respondent, Civil Action No. 20-mc-36-JEB-ZMF Facebook, Inc., Respondent. Civil Action No. 20-mc-36-JEB-ZM, p. 31.

⁸⁰ 18 U.S Code §2702 Voluntary disclosure of customer communications or records. Disponible en: 18 U.S. Code § 2702 - Voluntary disclosure of customer communications or records | U.S. Code | US Law | LII / Legal Information Institute (cornell.edu).

⁸¹ Esta decisión que parece que va en contra de la SCA, hizo que el propio tribunal tratara de limitar el alcance potencial de su fallo. En primer lugar, el tribunal enfatizó que solo se podía divulgar el contenido que se hubiera eliminado permanentemente de una plataforma en línea, no el material que había sido eliminado de la plataforma, pero aún no de manera permanente. En segundo lugar, el tribunal señaló que el contenido eliminado por el usuario de un servicio de comunicaciones electrónicas, a diferencia del contenido eliminado por el proveedor, no sería divulgable porque el contenido eliminado permanentemente por el usuario no era objeto de retención interna por parte de los proveedores parece no ser retenido por los proveedores. M. A. BECKER, “The Gambia v Facebook: Obtaining Evidence for Use at the International Court of Justice (Part I)”, *Blog of the European Journal of International Law*, 5th october 2021, que analiza el fallo: United States District Columbia, In re Application Pursuant to 28 U.S.C. § 1782 of The Republic of the Gambia, *op. cit.*, pp. 13-14. Disponible en: https://www.washingtonpost.com/context/order-in-re-the-republic-of-the-gambia-v-facebook-inc/6fd698bc-034f-43e2-a544-5592e174bc8a/?itid=lk_inline_manual_1.

⁸² 18 U.S Code §2702 (b)(3): “with the lawful consent of the originator or an addressee or intended recipient of such communication, or the subscriber in the case of remote computing service;”

⁸³ In re Application Pursuant to 28 U.S.C. § 1782 of The Republic of the Gambia, Petitioner, v. Facebook, Inc. *op. cit.*, pp.

público antes de que *Facebook* lo eliminara. De hecho, este material ya había sido proporcionado a la Misión Independiente de las NU durante los años en los que se investigaron los hechos en terreno y se prepararon los informes para el Consejo de Derechos Humanos⁸⁴.

51. Sin embargo, como ya hemos indicado con anterioridad, la clasificación del material publicado en las redes sociales como público o privado no siempre es sencilla. De hecho, el Tribunal quiso dar a entender que existe una compleja casuística – publicaciones de *Facebook* que están configuradas para ser públicas, abiertas para cualquier persona en cualquier momento; otras que son estrictamente privadas, o que se comparten solo con miembros de un grupo privado o seguidores de una página, que pueden requerir incluso de un administrador que otorgue acceso; y otras, que, sin embargo, aunque sean publicaciones privadas si están disponibles para el público en general u otorgan automáticamente acceso “serían efectivamente públicas”-. Ante estas situaciones, los tribunales deben realizar una investigación intensiva, basada en hechos, destinada a establecer si un usuario pretende que una comunicación sea de naturaleza pública o privada, aunque no sea tarea sencilla desvelar la intención subjetiva del usuario solo de factores objetivos⁸⁵.

52. Aplicando este análisis intensivo de los hechos, el Tribunal concluyó que, si bien parte del material buscado por Gambia se refería a páginas que eran nominalmente privadas y que pertenecían a funcionarios de Myanmar, los autores mostraron abiertamente su intención de llegar al público –llegaron a tener una audiencia de 12 millones de seguidores-. Como resultado, *Facebook* podría divulgar, sobre la base de la excepción de consentimiento legal recogida en la SCA, el material buscado por Gambia, salvo los mensajes privados de usuario a usuario que Gambia y *Facebook* acordaron que estaban más allá de esa excepción⁸⁶.

53. Esta decisión, por un lado, bienvenida, puede traer un efecto contraproducente porque si las empresas se ven abordadas judicialmente con peticiones masivas de divulgación de documentos eliminados, podrían decidir limitar unilateralmente esta práctica y, de esta manera, dificultar aún más la lucha contra el discurso del odio; esto si no llega antes una reforma legislativa, por otro lado necesaria, que vaya en la línea de abordar los conflictos entre los requisitos competitivos de privacidad y la protección de datos, la moderación de contenido y la libertad expresión, la cooperación con iniciativas de responsabilidad legal y la protección de derechos humanos.

54. Con relación a la segunda de las peticiones realizadas por Gambia, el acceso a los registros relacionados con la investigación interna de *Facebook*, el Tribunal acordó que estos esclarecerían cómo la empresa había obrado para conectar cuentas aparentemente falsas con funcionarios de Myanmar, y cómo esta conexión indujo a pensar que constituían una forma de propagación oculta y coordinada de odio anti-*rohingya* o de incitación a la violencia⁸⁷, por lo que terminaron siendo eliminadas. El Tribunal consideró que la información reclamada por Gambia era relevante para probar la intención genocida por parte de las autoridades de Myanmar y accedió a la solicitud de obligar a *Facebook* a presentar toda la documentación relacionada con su investigación interna, normalmente protegida bajo las normas del secreto profesional⁸⁸.

21. Es interesante, al menos, señalar que Gambia planteó una objeción preliminar basada en la tesis de que los funcionarios del gobierno de Myanmar no eran “usuarios protegidos” en virtud de la 18 U.S Code §2702 y, por lo tanto, no tenían derecho a la protección contra la divulgación. El Tribunal, sin embargo, señaló que todas las personas, incluidos los agentes de gobiernos extranjeros, tenían derecho a la protección contra el acceso no autorizado en virtud de la SCA. Véase en *The Gambia v. Facebook in Global Freedom of Expression*. Columbia University. Disponible en: <https://globalfreedomofexpression.columbia.edu/cases/gambia-v-facebook>.

⁸⁴ A/HRC/42/CRP.5 y A/HRC/39/CRP.5.

⁸⁵ In re Application Pursuant to 28 U.S.C. § 1782 of The Republic of the Gambia, Petitioner, v. Facebook, Inc., *op. cit.*, pp. 21-22.

⁸⁶ *Ibidem*.

⁸⁷ *Ídem*, p. 27.

⁸⁸ In re Application Pursuant to 28 U.S.C. § 1782 of the Republic of the Gambia, Petitioner, v. Facebook, Inc., *op. cit.*, pp. 28-29.

55. Dicho todo lo anterior, el interés de este causa reside en el hecho de que aunque el demandante, Gambia, no pretendiera establecer la responsabilidad de la empresa *Facebook* por permitir la divulgación de mensajes de incitación al odio y a la violencia, sí ha tratado de buscar la evidencia de que efectivamente estos mensajes se propagaron por la red y de que tuvieron una relación causal con los actos de violencia coordinada y sistemática que se produjeron contra los *rohingya*, lo que manifiesta una intención genocida por parte de Myanmar. Asimismo, cabe destacar que este caso ha sido funcional, no solo a los esfuerzos que Gambia está realizando para establecer la intención genocida de Myanmar ante la CIJ, sino para alimentar debates más amplios sobre las implicaciones que tiene el discurso de odio online sobre los límites a la libertad de expresión, la privacidad y la protección de datos, los derechos y obligaciones de los usuarios y, de manera particular, sobre la responsabilidad de las empresas de redes sociales en su difusión, cuestión que analizaremos a continuación.

V. De la responsabilidad civil en *Doe c. Meta* en Estados Unidos a la responsabilidad internacional de *Facebook* en el genocidio rohingya

56. Si en septiembre de 2021, el Tribunal Federal para el Distrito de Columbia ordenaba a *Facebook* entregar a Gambia la documentación requerida por su utilidad probatoria en el caso contra Myanmar en la CIJ, en el que se había alegado la violación de la Convención sobre el genocidio de 1948, en el mes de diciembre de ese mismo año, el despacho de abogados *Edelson & Fields Law* presentó una demanda colectiva ante el Tribunal Superior del Estado de California para el Condado de San Mateo contra *Meta Platforms*, la empresa matriz de *Facebook*, en nombre y representación de refugiados *rohingya* (*Doe c. Meta*).

57. La demanda, de acuerdo con el Código de Procedimiento Civil de California, Sección 382⁸⁹, tiene dos fundamentos legales: (i) la responsabilidad de la empresa por el producto –*strict product liability*– y (ii) negligencia –*negligence*–. Los demandantes han solicitado una compensación por un monto de al menos 150 mil millones de dólares por los daños causados sobre la población *rohingya*: las muertes producidas, las lesiones personales, el dolor, el sufrimiento, la angustia emocional causada y la pérdida de propiedad⁹⁰.

58. La demanda afirma que *Facebook* desempeñó un papel muy importante en la divulgación de mensajes de odio en Myanmar y en la violencia étnica desatada contra los *rohingya*, lo que se sostiene en los informes de la Misión Independiente de la ONU, informes de organizaciones de derechos humanos y de asociaciones internacionales de prensa, informes independientes que solicitó la empresa y en las declaraciones de sus altos cargos en las que se reconoció abiertamente que “no hicieron lo suficiente” para parar este posible genocidio⁹¹.

59. Respecto del primero de los fundamentos alegados, la responsabilidad por producto se refiere a la responsabilidad que tiene cualquiera o todas las partes que intervienen a lo largo de la cadena de fabricación de cualquier producto con defectos inherentes por los daños que provoque⁹². De hecho,

⁸⁹ California Code of Civil Procedure § 382. Disponible en <https://leginfo.ca.gov/faces/codesTOCSelected.xhtml?tocCode=CCP&tocTitle=+Code+of+Civil+Procedure+-+CCP>.

⁹⁰ Superior Court of the State of California for the County of San Mateo. *Jane Doe, individually and on behalf of all others similarly situated, Plaintiff, v. META PLATFORMS, INC.* (f/k/a Facebook, Inc.), a Delaware corporation, Defendant.

⁹¹ Jen Kirby, Mark Zuckerberg on Facebook’s role in ethnic cleansing in Myanmar: ‘It’s a real issue’, VOX, Apr. 2, 2018. Disponible en: <https://www.vox.com/2018/4/2/17183836/mark-zuckerbergfacebook-myanmar-rohingya-ethnic-cleansing-genocide>.; Kevin Roose, Paul Mozur, Zuckerberg Was Called Out Over Myanmar Violence. Here’s His Apology, NEW YORK TIMES, Apr. 9, 2018. Disponible en: <https://www.nytimes.com/2018/04/09/business/facebook-myanmar-zuckerberg.html>. On April 10, 2018, Zuckerberg testified before the U.S. Senate: Facebook, Social Media Privacy, and the Use and Abuse of Data, Senate Hearing 115- 683 before the Comm. on Commerce, Science, and Transportation, et al., 115th Cong. (Apr. 10, 2018). Disponible en: <https://www.govinfo.gov/content/pkg/CHRG-115shrg37801/html/CHRG-115shrg37801.htm>.

⁹² Véase su significado en https://www.law.cornell.edu/wex/products_liability.

este es uno de los puntos fuertes de esta demanda ya que, en línea con el caso *Lemmon v. Snap, Inc.*⁹³, se apela a la responsabilidad de *Meta*, no como distribuidor o editor de contenidos, sino como fabricante de producto. En este sentido, los demandantes han señalado básicamente que el sistema de *Facebook* y los algoritmos subyacentes están programados para maximizar la participación de usuarios porque cuanto mayor sea su número, mayores serán los ingresos publicitarios y, por consiguiente, sus beneficios⁹⁴. De hecho, es interesante subrayar que los demandantes han argumentado que *Facebook* conoce muy bien cómo las emociones más negativas, tales como el miedo o el odio, resultan ser las más atractivas para los usuarios.⁹⁵ De esta forma, para maximizar la participación de los miembros de la red, *Facebook* no solo llena las fuentes de noticias de los usuarios con cantidades desproporcionadas de mensajes de incitación al odio y desinformación a través de sus algoritmos, sino que además emplea un sistema de recompensas que manipula y entrena a los usuarios para que creen dicho contenido⁹⁶; esto explica, con mucha nitidez, cómo los algoritmos y el sistema que emplea *Facebook* facilitarían, por sí mismos, que los mensajes de odio hacia los *rohingya* incendiaran las redes sociales⁹⁷. Sobre la base de este argumento, los demandantes han sostenido que *Meta* sería responsable de crear un algoritmo que “lejos de ser neutral” respecto de los mensajes de odio, habría creado un “mecanismo de reacción desencadenante de dopamina” que intencionalmente promovió el contenido de odio con respecto de los *rohingya*. De hecho, parece que los demandantes han querido anticiparse a una respuesta del Tribunal en línea con la del caso *Gonzales v. Google LLC*⁹⁸, en el que se sostuvo que los algoritmos de las empresas tecnológicas demandadas, *Google, Facebook, Twitter*, eran “de contenido neutral” en relación con la actividad terrorista del ISIS y tenían derecho a la inmunidad que otorga la Ley de Decencia en las Comunicaciones Sección 230 (*Communications Decency Acts* §230, a partir de ahora Ley CDA-sección 230) que es una de las herramientas legales más valiosas para proteger la libertad de expresión y la innovación en internet en los Estados Unidos⁹⁹.

60. Los demandantes, además, respecto de la cuestión de la responsabilidad por producto, han añadido una consideración que es de enorme relevancia a efectos de establecer una relación causal entre el fallo del sistema y los daños: el tipo de daño resultante de la violencia étnica cometido contra los *rohingya* - muertes, lesiones, violaciones, daños psicológicos y morales y pérdida de la propiedad- consistía precisamente en el tipo de daño que podría razonablemente haberse esperado de la incitación al odio que se llevó a cabo a través de la red social¹⁰⁰.

⁹³ El Noveno Circuito consideró recientemente el mismo argumento en una demanda contra Snapchat. En *Lemmon v. Snap, Inc.*, los demandantes alegaron que Snapchat diseñó negligentemente un “filtro” que incentivaba a los usuarios a enviar fotos mientras conducían vehículos a altas velocidades. El Noveno Circuito sostuvo que la inmunidad de la Sección 230 no se aplicaba porque los demandantes buscaban responsabilizar a Snapchat por su deber específico de diseñar un producto razonablemente seguro, en lugar de buscar responsabilizar a Snapchat como editor.

⁹⁴ *Jane Doe, individually and on behalf of all others similarly situated, Plaintiff, v. META PLATFORMS, INC, op. cit.*, párrs. 37 y 167. Hace referencia a L. MUNN, “Angry by design: toxic communication and technical architectures”, *Humanit Soc Sci Commun* 7 July 30, 2020. Disponible en: <https://www.nature.com/articles/s41599-020-00550-7>.

⁹⁵ *Ídem*, párr. 41.

⁹⁶ *Ídem*, párrs. 42, 167, 168 y 169.

⁹⁷ Como ya se ha hecho referencia anteriormente, y tal y como está reportado en los informes de la Misión Independiente de Investigación de la ONU, *Facebook* gozaba de una posición privilegiada en Myanmar con motivo de los acuerdos suscritos con gobierno y empresas. A/HRC/42/CRP.5 y A/HRC/39/CRP.5.

⁹⁸ En *Gonzales v. Google LLC*, demandaron a *Google, Facebook y Twitter*, alegando que estas empresas de tecnología facilitaron las actividades terroristas de ISIS al recomendar videos de reclutamiento de ISIS a los usuarios y permitirles ubicar otros videos y cuentas relacionadas con ISIS. El Noveno Circuito encontró que los demandantes no habían alegado (1) que los algoritmos de *Google* habían incitado a ISIS a publicar contenido ilegal, o (2) que los algoritmos de *Google* trataron el contenido creado por ISIS de manera diferente a cualquier otro contenido creado por terceros. Por lo tanto, la Corte de Apelaciones del Noveno Circuito sostuvo que los algoritmos de estas empresas eran “de contenido neutral” y tenían derecho a la inmunidad de la Sección 230. *Reynaldo Gonzalez, et al., Petitioners, United States Court of Appeals for the Ninth Circuit, Case Numbers:(18-16700)*. El caso puede seguirse en: <https://www.supremecourt.gov/docket/docketfiles/html/public/21-1333.html>. Sobre la decisión de la Corte de Apelaciones puede leerse en: *Gonzalez v. Google LLC* en <https://www.law.cornell.edu/supct/cert/21-1333> [última consulta, 14 de marzo de 2023].

⁹⁹ Y, por lo tanto, de proteger de cualquier responsabilidad civil a las plataformas de servicios online del contenido que crean, produce y emiten terceros a través de estas plataformas. La Sección 230 puede leerse en: <https://www.law.cornell.edu/uscode/text/47/230>.

¹⁰⁰ *Ídem*, párr. 170.

61. A propósito del segundo de los fundamentos de la demanda, la negligencia, los demandantes han argumentado que *Facebook* tenía un deber de cuidado que incumplió, ya que gozando de una posición privilegiada en Myanmar con motivo de los acuerdos suscritos con gobierno y empresa¹⁰¹, no contrató suficientes moderadores en el idioma birmano¹⁰² y respondió de una forma muy lenta frente a las diferentes informaciones que organizaciones y particulares hacían llegar denunciando los discursos de odio que incendiaron *Facebook*¹⁰³. Para los demandantes, los acciones y omisiones de *Facebook*, conectadas con el incumplimiento de su deber de cuidado, fueron determinantes para la persecución y la violencia étnica en contra de los *rohingya*¹⁰⁴. Ahora bien, aunque los argumentos por negligencia son sólidos, cabe señalar que aquí la teoría legal de la responsabilidad se basa en los actos y las omisiones de *Meta* como moderador de contenido de terceros, lo que se aleja de la estrategia legal desarrollada respecto del primero de los argumentos -fabricante de producto- y esto puede ser un problema de cara al establecimiento de la responsabilidad en el caso si la intención de los demandantes era evitar la aplicación de la inmunidad que ofrece la CDA-230 .

62. De hecho, *Facebook* ha respondido a esta demanda señalando que no es responsable por lo que los usuarios comunican en la red social, amparada bajo la CDA-230¹⁰⁵. Bajo esta norma, los intermediarios de plataformas online- que incluyen no solo proveedores de servicios de Internet (ISP) regulares, sino también una variedad de proveedores de servicios informáticos interactivos, que incluyen básicamente cualquier servicio en línea que publique contenido de terceros- están protegidos legalmente frente a lo que otros dicen y hacen. Así, ha permitido que empresas como *Facebook*, que tiene más de 1000 millones de usuarios, no deban enfrentarse a una posible responsabilidad legal por las acciones de sus usuarios. Si no fuera así, la mayoría de estas plataformas no alojarían contenidos o tendrían que protegerse participando activamente en la censura de lo que decimos, vemos y hacemos en línea¹⁰⁶. Ahora bien, la inmunidad que ofrece la aplicación de la Ley CDA-230 contiene algunas excepciones que operan en casos en los que las actividades online indujeran o alentaran al desarrollo de contenido ilegal, así como cuando no se advirtiera a los usuarios sobre actividades ilegales, incumpliera un contrato y también si no actuara de buena fe¹⁰⁷.

63. La demanda *Doe c. Meta* ha sido desestimada, en diciembre de 2022, por la jueza del Tribunal Federal para el Distrito Norte de California, División de San Francisco, en el que se encuentra el caso después de la petición de traslado que hicieron los demandados¹⁰⁸. En la decisión judicial, la jueza Yvonne Gonzalez Rogers rechazó los cargos de negligencia y responsabilidad por producto defectuoso alegados por el demandante. De acuerdo con las fuentes consultadas¹⁰⁹, la juez indicó, por un lado, que la demanda no establecía con claridad la relación causal entre la supuesta lesión que sufrió la demandante -un ataque a su aldea por parte del ejército de Myanmar- y la plataforma de *Facebook*. Además, por otro lado, razonó que *Facebook* tampoco se define como un producto en California, por lo que el demandante

¹⁰¹ A/HRC/42/CRP.5 y A/HRC/39/CRP.5. la gran mayoría de los ciudadanos birmanos obtuvieron teléfonos celulares en 2011 y que Facebook hizo arreglos para que usaran la aplicación de Facebook sin incurrir en cargos por datos. Para muchos birmanos, Facebook era Internet, lo que facilitaba mucho la difusión de retórica de odio.

¹⁰² *Jane Doe, individually and on behalf of all others similarly situated, Plaintiff, v. META PLATFORMS, INC. op. cit.*, párr.145.

¹⁰³ Ídem, párr.173.

¹⁰⁴ Ídem, párrs. 174-180.

¹⁰⁵ 230 Communications Decency Acts (1996). Disponible en: <https://www.congress.gov/bill/104th-congress/senate-bill/314/>.

¹⁰⁶ Ídem. De hecho, existen precedentes en los que el Noveno Circuito ha sostenido que las empresas de tecnología como Facebook, Twitter y Google están inmunizadas de responsabilidad civil cuando actúan como “editores” de información publicada por terceros.

¹⁰⁷ Información disponible en: <https://itif.org/publications/2021/02/22/exceptions-section-230-how-have-courts-interpreted-section-230/>.

¹⁰⁸ Case No. 4:22-cv-00051-ygr, Motion by defendant meta platforms, inc. To dismiss complaint pursuant to rule 12(b)(6).

¹⁰⁹ Hay que tener en cuenta que al momento de escribirse este artículo la decisión de la jueza Yvonne Gonzalez Rogers del Tribunal Federal para el Distrito Norte de California, División de San Francisco, no se encuentra en abierto y no se ha podido estudiar su contenido. Por esa razón, la información en este artículo se desprende de una fuente periodística: <https://observer.com/2022/12/the-rohingyas-genocide-suit-against-meta-is-dismissed-for-now/>

no podía reclamar responsabilidad por defecto en el producto de acuerdo con el Código de Procedimiento Civil de California, Sección 382. Para concluir, además, la jueza afirmó que la Ley CDA-230 otorga inmunidad a plataformas como *Facebook* que publican contenido de terceros¹¹⁰.

65. Aunque la desestimación del caso por el Tribunal Federal pueda parecer un importante revés, deja la puerta abierta para que los demandantes vuelva a presentar el caso, y puede ser una excelente ocasión para que reformulen la acusación y que, para ese entonces, un tribunal se pueda pronunciar sobre cuestiones tan relevantes como si el diseño del sistema de *Facebook* tiene defectos que dañaron a los *rohingya* y si la empresa es responsable por ello; si la empresa contrajo un deber de cuidado hacia la población cuando entró en el mercado de Myanmar; si rompió ese deber de cuidado de acuerdo con la forma en la que operó en el país; y, finalmente, si las operaciones de *Facebook* causaron un daño a los *rohingya*¹¹¹.

66. Además, se espera que pueda responder a la cuestión de si *Facebook* goza de inmunidad en virtud de la Ley CDA-230 o si las actividades en la plataforma indujeron a los actos de violencia o la comisión de un genocidio contra los *rohingya* operando una excepción a la inmunidad que ofrece¹¹². Sobre esta última cuestión, hay que señalar que cualquier decisión que se adopte es de enorme trascendencia. De hecho, este caso es un ejemplo excelente de cómo han cambiado los tiempos y de cómo la realidad empresarial actual de las empresas tecnológicas, que juegan un papel vital en la difusión de información, difiere notablemente de la realidad que existía en 1996 cuando se adoptó la Ley CDA -230. Es así como el caso *Doe c. Meta* es un claro recordatorio de la necesidad de reformar la Sección 230 porque las empresas de tecnología como *Meta* ejercen una enorme influencia en la difusión de información, particularmente en países en desarrollo con bajos índices de alfabetización digital, y los algoritmos de aprendizaje automático pueden facilitar la violencia, la desinformación y los crímenes más atroces¹¹³.

67. Por lo tanto, las decisiones que se adopten dentro de este caso serán trascendentales en la medida que colocarán la balanza hacia la protección de la innovación y la libertad de expresión online, de acuerdo con una larga tradición bajo la Primera Enmienda de la Constitución de los Estados Unidos, o hacia una protección más amplia de los derechos humanos que se ve interpelada cuando una empresa global permite que crímenes atroces puedan cometerse solo para garantizar sus beneficios, como es el caso de *Facebook*. Asimismo, no serán relevantes sólo para los Estados Unidos, sino que tendrán una repercusión global en la medida que es en este Estado en el que se han establecido las más importantes empresas de comunicación e intermediarios online, con miles de millones de usuarios en todo el mundo; si las grandes empresas intermediarias online están responsabilizan o estuvieran obligadas a responsabilizarse de garantizar de manera efectiva que sus plataformas no se utilizan para incitar al odio, el impacto de un posible control legal tendría una repercusión global¹¹⁴.

68. Ante esta tesis, se plantea a través de qué mecanismos normativos las empresas pueden hacerse responsables o ser responsabilizadas por los discursos de odio online que se crean, divulgan y promueven en sus plataformas online.

69. Por un lado, los mecanismos de autorregulación se han considerado por largo tiempo claves para que las empresas aborden los riesgos de sus actividades cuando no referimos al campo de los

¹¹⁰ *Ídem*.

¹¹¹ *Jane Doe, individually and on behalf of all others similarly situated, Plaintiff, v. META PLATFORMS, INC. op. cit.*, párr. 161.

¹¹² Lo cierto es que la norma se ha vuelto cada vez más controvertido a medida que crece el poder de las empresas de Internet. Véase, por ejemplo: <https://www.jonesday.com/en/insights/2021/04/section-230-a-focus-of-criticism-reform-proposals>.

¹¹³ De hecho, existe un consenso de que el sistema actual de autorregulación está obsoleto y para abordar estos problemas el Congreso debería enmendar la Sección 230 de la CDA. Véase A. PREVE, *Doe vs. Meta and the Future of the Communications Decency Act*, Transnational Litigation Blog, May 25, 2022.

¹¹⁴ Queremos señalar que no se ha abordado la cuestión de la presunción contra la extraterritorialidad de la aplicación de la Ley CDA-230 pues supera el alcance de este trabajo.

derechos humanos que, por motivos obvios, está íntimamente vinculado con los daños que produce el discurso del odio que llama a la violencia y a la discriminación étnica o racial. El informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, *David Kaye*, presentado en el año 2019, recomienda a las empresas que implementen procesos de debida diligencia en la moderación de los contenidos que pueden ser discursos del odio y que lo hagan además desde un enfoque de derechos humanos y, por tanto, en línea con Principios Rectores sobre Empresas y Derechos Humanos de la ONU¹¹⁵. Para ello invita a las empresas a que vayan más allá de las normas de la comunidad y de los términos de referencia de sus contratos y evalúen de manera periódica y pública la forma en la que sus productos y servicios afectan a los derechos humanos de sus usuarios y al público en general. Asimismo, apela a que adopten políticas sobre el contenido de sus actividades y que vinculen directamente sus normas sobre el discurso de odio con el Derecho internacional de los derechos humanos. Además, propone que las empresas definan 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 y que velen por que la aplicación de las normas relativas al discurso de odio entrañe una evaluación del contexto y del daño, en particular, garantizando que el empleo de instrumentos automáticos o de inteligencia artificial incluya a un ser humano informado. El Relator Especial llama a que velen por la participación de las comunidades afectadas en el análisis del contenido y en la búsqueda de instrumentos eficaces para hacer frente a los daños que se generan desde las plataformas. Por último, propone que elaboren instrumentos que además de promover la libre expresión libre, incluyan medidas para educar, contrarrestar, informar y capacitar frente al discurso del odio¹¹⁶.

70. A la autorregulación ha contribuido el *soft law* en términos de recomendaciones, guías y directrices que orientan a las empresas sobre cómo abordar esta tarea. En particular, se han producido algunos desarrollos interesantes en el marco europeo. La Comisión Europea ha emitido la Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico Social Europeo y al Comité de las Regiones: abordar el contenido ilegal en línea hacia una mayor responsabilidad de las plataformas en línea de 2017¹¹⁷. Además, la Comisión adoptó en 2018 la Recomendación (UE) 2018/334 de la Comisión sobre medidas para combatir eficazmente los contenidos ilícitos en línea¹¹⁸. Finalmente, cabe mencionar el Código de Conducta de la UE relativo a la lucha contra la incitación ilegal al odio en internet de 2016, adoptado por cuatro grandes plataformas de Internet (*Facebook*, *YouTube*, *Twitter* y *Microsoft*) en junio de 2016¹¹⁹ y que otras cuatro plataformas de Internet más (*Google +*, *Instagram*, *Snapchat* y *Daily Motion*) acordaron implementar¹²⁰. De acuerdo con los datos arrojados por el informe de evaluación del Código de Conducta sobre el estado de la cuestión de 2019¹²¹, y para continuar poniendo en el centro a la actividad de la empresa *Facebook*, ésta informó lo siguiente: tener a 15.000 personas trabajando en la revisión de

¹¹⁵ Consejo de Derechos Humanos. Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, A/47/486, 9 octubre 2019.

¹¹⁶ Ídem, párr. 58.

¹¹⁷ Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones: Lucha contra el contenido ilícito en línea. Hacia una mayor responsabilización de las plataformas en línea COM (2017) 555 final, Bruselas, 28.09.2017. La Comunicación establecía una serie de directrices para garantizar que los intermediarios "intensifican" la lucha contra el contenido ilegal en línea

¹¹⁸ Recomendación (UE) 2018/334 de la Comisión, de 1 de marzo de 2018, sobre medidas para combatir eficazmente los contenidos ilícitos en línea. La recomendación propone que los intermediarios de Internet establezcan mecanismos de notificación y acción en sus sitios web mediante los cuales los usuarios pueden llamar la atención de los intermediarios sobre el contenido ilegal en línea para que los intermediarios pueden eliminar o bloquear el acceso a contenido que es ilegal. Disponible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32018H0334&from=EN>.

¹¹⁹ European Commission. Assessment of the Code of Conduct on Hate Speech online State of Play. Disponible en: https://ec.europa.eu/info/sites/default/files/aid_development_cooperation_fundamental_rights/assessment_of_the_code_of_conduct_on_hate_speech_on_line_-_state_of_play_0.pdf

¹²⁰ European Commission, Code of Conduct on countering illegal hate speech online: Results of the 3rd monitoring exercise (Fact sheet, January 2018)8). Disponible en: http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=612086.

¹²¹ European Commission. Assessment of the Code of Conduct on Hate Speech online State of Play. Disponible en: https://ec.europa.eu/info/sites/default/files/aid_development_cooperation_fundamental_rights/assessment_of_the_code_of_conduct_on_hate_speech_on_line_-_state_of_play_0.pdf.

todo tipo de contenidos que podían violar las políticas de la compañía¹²²; haber establecido un Foro sobre Política de Producto en el que reúne a todos sus expertos en todo el mundo cada dos semanas, para discutir posibles cambios en las normas de la comunidad y abordar nuevos problemas, tendencias y desarrollos¹²³; haber incrementado el número de medidas adoptadas por violaciones del discurso de odio aumentando hasta un 150% - 4 millones de acciones tomadas contra el discurso del odio¹²⁴; y haber presentado informes de transparencia y en los que dedicaba un espacio para hacer cumplir los estándares de la comunidad respecto del discurso de odio¹²⁵, por mencionar alguno de los avances que fueron informados.

71. Lo cierto es, por tanto, que las empresas tienen un enorme margen para establecer restricciones a las formas de expresión de odio en el conjunto de las normas internas de la comunidad y, siguiendo algunas de estas directrices y recomendaciones, se han ido produciendo algunos avances, en particular, entre los proveedores de servicios más importantes como *Facebook*, *Twitter* o *Youtube*¹²⁶.

72. Sin embargo, estas normas de autorregulación o de *soft law*, basadas en la debida diligencia en derechos humanos, no son suficientes para abordar los desafíos que impone el discurso online que incita al odio¹²⁷. A pesar de que *Facebook* haya mejorado aspectos relativos al seguimiento e identificación del discurso del odio, la tecnología para identificar ciertos tipos de contenido que violan sus estándares, facilitada para que los usuarios puedan reportar situaciones que deben ser revidas e incluso eliminadas por los moderadores, no parece haber sido suficiente para detener el odio contra los *rohingya* en Myanmar, menos aún para hacer surgir algún tipo de responsabilidad legal de la empresa por estos hechos.

73. Por esta razón, es necesario y urgente que los Estados definan en su Derecho interno, con claridad y siguiendo los estándares internacionales, qué es el discurso de odio punible y, más importante si cabe, qué hacer frente a los retos que presenta el discurso de odio online hoy en día. De esta forma, es importante que los Estados examinen el derecho vigente o elaboren leyes sobre el discurso de odio que cumplan los requisitos de legalidad, necesidad, proporcionalidad y legitimidad y que aprueben o revisen las normas relativas a la responsabilidad del intermediario, aunque no puedan exigir que las empresas restrinjan expresiones que los Estados no puedan hacer directamente mediante leyes. Además, estos deben aprobar leyes que requieran que las empresas describan en detalle y en público el modo en que definen el discurso de odio y que apliquen sus normas en contra de él y que creen bases de datos de las medidas adoptadas para combatirlo. De esta forma se alienta también a las empresas a respetar las normas de derechos humanos en sus propias reglas¹²⁸. En definitiva, el propósito es avanzar en el desarrollo de un marco normativo de base nacional, lo suficientemente robusto, para luchar contra el discurso de odio y los medios que lo difunden a gran escala desde espacios fuera de cualquier control normativo y jurisdiccional.

74. De la misma manera, es necesario avanzar en el Derecho internacional hacia la creación de obligaciones de los proveedores de servicios de internet en el control del discurso del odio y en el establecimiento de algún tipo de responsabilidad cuando este discurso se divulgue a través de las plataformas. Los desarrollos normativos que existen hasta el momento que abordan este tema, aunque sea tangencialmente, no son suficientes. Así, por ejemplo, la Directiva 2000/31 sobre comercio electrónico¹²⁹, que

¹²² *Ídem*, p. 3.

¹²³ *Ídem*, p. 4.

¹²⁴ *Ibidem*.

¹²⁵ *Ídem*, p. 5.

¹²⁶ Disponibles en: <https://support.google.com/youtube/answer/2801939?hl=en>, www.facebook.com/communitystandards/hate_speech and <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>.

¹²⁷ A. GUAMÁN HERNÁNDEZ / G. MORENO GONZÁLEZ, *Empresas Transnacionales y Derechos Humanos. La necesidad de un Instrumento Vinculante*, Editorial Bomarzo, Albacete, 2018; C. MONTESINOS PADILLA, *Los derechos humanos en el ámbito empresarial: qué, cómo y por qué es posible regular en la materia*, Tirant lo Blanch, 2020.

¹²⁸ A/47/486, *op. cit.* párr. 57.

¹²⁹ Directiva 2000/31/CE del Parlamento Europeo y del Consejo, de 8 de junio de 2000, relativa a determinados aspectos jurídicos de los servicios de la sociedad de la información, en particular el comercio electrónico en el mercado interior (Directiva sobre el comercio electrónico). Disponible en <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32000L0031>.

debe ser traspuesta al derecho interno de los Estados miembro, dispone que los intermediarios de internet quedarán exentos de cualquier responsabilidad por los contenidos en sus plataformas cuando un intermediario se dé cuenta de que está alojando contenido ilegal y lo elimine de manera expedita o deshabilite el acceso a la misma¹³⁰; y, además, cuando no hayan generado o aprobado el contenido¹³¹. La Directiva también establece que los Estados miembro no pueden imponer una obligación general a los intermediarios de monitorear todo el contenido que manejan, lo que restringe los mecanismos a disposición de los mismos Estados en relación con la regulación de contenido¹³² y, por tanto, del establecimiento de responsabilidad. Sin embargo, esto no afecta al hecho de que los Estados puedan exigir a los prestadores de servicios que apliquen un deber de diligencia – al menos una expectativa de comportamiento razonable de acuerdo con el Derecho interno, a fin de prevenir y evitar determinados tipos de actividades ilegales¹³³

75. De otro lado, la Decisión Marco 2008/913/JAI del Consejo, de 28 de noviembre de 2008, relativa a la lucha contra determinadas formas y manifestaciones de racismo y xenofobia mediante el derecho penal, sí establece la responsabilidad penal de las personas físicas y de las personas jurídicas por las conductas prohibidas¹³⁴. Ahora bien, respecto del discurso de odio online, que no está explícitamente contemplado en la norma, habrá que tener en cuenta dos circunstancias que podrán dificultar su aplicación a los proveedores de servicios: de un lado, la persona jurídica será considerada responsable cuando las conductas prohibidas sean realizadas solo en su provecho por cualquier persona, actuando a título individual o como parte de un órgano de la persona jurídica, que ostente un cargo directivo en el seno de dicha persona jurídica¹³⁵; de otro lado, cuando la falta de vigilancia o control por parte de una de las personas referidas previamente, haya hecho posible que una persona sometida a la autoridad de la persona jurídica realice algunas de las conductas contempladas, en provecho de esta última¹³⁶. Este estándar centrado en la idea del “provecho” para la atribución de responsabilidad supondrá una limitación respecto de los intermediarios de servicios del discurso de odio en la medida que solo demostrando que existe una relación directa entre las ganancias de las empresas y la difusión del odio online -como ya se ha mencionado con anterioridad- podría hacerla surgir.

76. Finalmente, tendremos que ver cuál es el texto definitivo de la Directiva del Parlamento Europeo y del Consejo sobre diligencia debida de las empresas en materia de sostenibilidad, presentada por la Comisión Europea, el 23 de febrero de 2022, y consecuentemente su alcance sobre las empresas proveedoras de servicios online.

77. Puesto que los desarrollos normativos actuales son inadecuados o insuficientes para tratar la responsabilidad de las empresas proveedoras de servicios o intermediarias por la difusión de mensajes de odio online que incitan a la violencia o la discriminación, surge la pregunta, en relación con el caso aquí analizado de la contribución de la empresa Facebook en el *presunto* genocidio *ronhingya*, de si no es posible tener a la empresa, al menos, como cómplice por la incitación al genocidio.

78. Ya se ha expuesto, en la primera parte de este trabajo, cómo algunos medios de comunicación han estado implicados en la comisión de genocidios y personas individuales han sido declaradas culpables del delito de incitación a este crimen de acuerdo con la Convención sobre el genocidio de 1948 y otros instrumentos jurídicos internacionales. Claro está que se ha hablado, en los casos referidos con anterioridad, de medios de comunicación tradicionales, en los que la identificación de los estándares sobre la incitación al odio ha sido más fácilmente practicable. Sin embargo, es razonable pensar que una

¹³⁰ A/47/486, *op. cit.*, párr. 46.

¹³¹ *Ídem*, párr. 43.

¹³² C. O'REGAN, “Hate Speech Online: an (Intractable)...”, *op. cit.*, p. 421.

¹³³ Decisión Marco 2008/913/JAI del Consejo, de 28 de noviembre de 2008, relativa a la lucha contra determinadas formas y manifestaciones de racismo y xenofobia mediante el derecho penal, párr. 48.

¹³⁴ *Ídem*, art. 3. Sanciones penales y art. 5. Responsabilidad de las personas jurídicas.

¹³⁵ *Ídem*, art. 5.1.

¹³⁶ *Ídem*, art. 5.2.

empresa proveedora de servicios de internet, como *Facebook*, pueda estar involucrada en un genocidio – como el de los *rohingya*- al haber permitido la incitación a este aberrante crimen, a través de su plataforma, con un diseño defectuoso del sistema y de sus algoritmos subyacentes y una respuesta negligente, todo ello movido por un ánimo de lucro ¹³⁷.

79. Si bien el Derecho internacional penal tiene como uno de sus principios centrales el de la responsabilidad internacional del individuo -y existen de hecho mecanismos para que pudiera dirimirse la responsabilidad de los responsables de una empresa, en particular, por medio de mecanismos judiciales nacionales y mecanismos internacionales *ad hoc*¹³⁸, no tanto el de la Corte Penal Internacional en la que no hay práctica al respecto¹³⁹- una cuestión distinta es la de la responsabilidad de las personas jurídicas, como son las empresas. Mucho se ha escrito y debatido sobre esta cuestión¹⁴⁰ pero no hay avances normativos significativos¹⁴¹. Entre los autores que han escrito sobre esta materia hay un cierto consenso en la idea de que la sola persecución individual de los miembros de la empresa no se ajusta a la dimensión colectiva del comportamiento criminal empresarial, y tampoco es suficiente para prevenirlo¹⁴². La consideración de la responsabilidad penal de las personas jurídicas, de hecho, empieza a gozar de una cierta preferencia en cuanto que solo así es posible modificar la estructura empresarial que facilita o promueve la comisión del delito; el proceso penal y la pena crea una mayor estigmatización en comparación con los procesos civiles y tiene un mejor efecto comunicativo y preventivo; y, además, la reparación se aprecia posible ya que las empresas suelen tener un patrimonio considerable¹⁴³.

80. El caso de la participación de *Facebook* en el genocidio *rohingya* es claramente un caso en el que los graves abusos del poder económico, la anteposición de los beneficios económicos al respeto de los derechos humanos- la incitación al odio ha sido un simple costo a la hora de hacer negocios- suceden cuando una empresa poderosa ejerce ese poder sin control sobre el mundo; y esto constituye un injusto merecedor de pena.

81. Para concluir, esto nos lleva a recordar algunas propuestas que se han elaborado para abordar el tipo de desafíos crecientes que para el Derecho internacional presenta la relación entre los crímenes internacionales y los actores económicos: los crímenes económicos contra la humanidad¹⁴⁴.

82. Cualquiera que sea la fórmula empleada en un caso como el de *Facebook*, la aplicación de normas de *lege lata* o la consideración de una propuesta de *lege ferenda*, habrá que tener siempre puesta la mirada en la reparación de las víctimas y en el fin de la impunidad.

¹³⁷ Como ya se indicó en el epígrafe anterior, *Facebook* habría mostrado más mensajes de odio, violencia y discriminación solo para alcanzar un mayor número de usuarios. *Jane Doe, individually and on behalf of all others similarly situated, Plaintiff, v. META PLATFORMS, INC.* *op. cit.*, párr. 52.

¹³⁸ *Special Tribunal for Lebanon (STL), Al Jadeed S.A.L. & Ms Khayat (STL-14-05)*.

¹³⁹ El anuncio realizado por el ex Fiscal de la Corte Penal Internacional, Luis Moreno Ocampo, en el que señalaba que: “[...] the Prosecutor believes that investigation on the financial aspects of the alleged atrocities will be crucial to prevent future crimes and for the prosecutions of crimes already committed” (sobre la situación en la República Democrática del Congo). Véase en ICC/OTP, Communications received by the Office of the Prosecutor of the ICC, Press release n°. pids.009.2003_EN, 16.7.2003.

¹⁴⁰ Se puede encontrar un resumen de las aportaciones doctrinales más importantes en L. ARENAL LORA, “La comisión de crímenes contra la humanidad por empresas transnacionales y la ampliación de la competencia *ratione personae* de la Corte Penal Internacional” *Deusto Journal of Human Rights*, Oct.: 1-29. DOI: <http://dx.doi.org/10.18543/djhr.1872>. K. AMBOS, *Derecho Penal Internacional Económico. Fundamentos de la responsabilidad penal internacional de las empresas*. Cuadernos Civitas. Thomson Reuters, 2018.

¹⁴¹ El Protocolo de Reforma del Protocolo del Estatuto de la Corte Africana de Justicia y Derechos Humanos o Protocolo de Malabo de 2014 es el primer instrumento estatutario que incorpora la responsabilidad penal directa de las personas jurídicas mediante el modelo de imputación basado en el modelo de organización.

¹⁴² L. ARENAL LORA, “La comisión de crímenes contra la humanidad...”, *op. cit.*

¹⁴³ K. AMBOS, *Derecho Penal Internacional Económico...*, *op. cit.*

¹⁴⁴ L. ARENAL LORA, *Crímenes económicos contra la humanidad: propuesta de una categoría de crímenes contra la humanidad*, Thomson Reuters-Aranzadi, 2019.

VI. Conclusiones

82. La relación entre el discurso del odio y la comisión de crímenes de trascendencia internacional es una realidad que ha podido constatararse en el marco de conflictos armados de alto impacto y en contextos de violencia generalizada y sistemática a lo largo de nuestra historia reciente. En particular, la incitación al genocidio fue tipificada en la Convención sobre el Genocidio de 1948, después de los aberrantes crímenes que se cometieron durante la Segunda Guerra Mundial.

83. Ahora bien, si en los casos en los que las palabras que incitan al odio, a la violencia o a la discriminación son punibles y se publican y se divulgan a través de medios de tradicionales – prensa, radio, televisión- la tarea de exigir responsabilidad legal por estos delitos es compleja, lo es mucho más como cuando el discurso del odio utiliza medios de comunicación online.

84. Estos medios presentan características propias y su utilización masiva y global presenta problemas particulares en este terreno: la identificación de los autores, en particular cuando varios actores participan en la creación y en la difusión del discurso del odio y se utilizan cuentas falsas; el establecimiento de la intención para lo que hay que tener en cuenta los rasgos de desinhibición, libertad, automatismo de la respuesta y de la sensación de anonimato que producen las especificidades del medio; la lejana posibilidad de la creación de un riesgo inminente debido a la fugacidad, la itinerancia y lo dispersa o difusa que es la audiencia, lo que imbrica, además, con la imprecisión del límite en la calificación del ámbito privado o público de estos mensajes; el *fórum-shopping* y los conflictos de aplicación de leyes y de jurisdicción; y, de manera especial, el papel de las grandes empresas que suministran servicios de internet, que actúan en la búsqueda permanente del lucro, y que no tienen una obligación general de controlar el contenido que circula en sus redes.

85. De esta manera, todos y cada uno de estos elementos constituyen un enorme desafío en materia de identificación del discurso de odio online punible y de la persecución de estos delitos y de sus responsables. Sin embargo, de todos ellos, exige particular atención el de la responsabilidad de las empresas proveedoras de servicios, empresas que, por otro lado, tienen un enorme poder que ejercen globalmente.

86. Las empresas cuentan con un enorme margen para autorregularse y establecer límites propios en la divulgación de los discursos del odio. Para abordar esta tarea cuentan con la orientación de guías y directrices internacionales debido a que, en los últimos tiempos, el Derecho internacional ha realizado una importante labor para promover que las empresas lleven a cabo sus actividades siguiendo normas de debida diligencia en materia de derechos humanos. Ahora bien, la práctica también ha mostrado que estos mecanismos de *soft law* no son suficientes por sí mismo y, por tanto, hay que acudir a normas que creen obligaciones para las empresas cuando la actividad económica participa de una forma sustancial en la comisión de delitos o de crímenes de trascendencia internacional. Esta será la única forma en la que pueda exigirse responsabilidad legal –la de las personas individuales que dirigen y controlan las empresas y la de las personas jurídicas-. Uno de los campos que pueden explorarse en este sentido es el del desarrollo progresivo de normas del Derecho internacional penal en donde las empresas podrían reconocer mejor las expectativas normativas que le son dirigidas y ajustar su política empresarial, además de conocer las consecuencias de su responsabilidad.

87. La resolución de las diferentes demandas judiciales en el caso del genocidio contra los *rohingya* en el que está involucrada la empresa *Facebook* serán, sin lugar a duda, relevantes para comprobar si se están produciendo avances significativos en la tarea de exigir responsabilidad a las empresas por sus comportamientos económicos criminales y cómo se pueden superar las barreras de la extraterritorialidad.