

DISCURSO EXTREMISTA Y LA EXPANSIÓN DE LA CENSURA¹

CÉSAR FRANCISCO GALLEGOS PAZMIÑO²
EDDY ISMAEL VILLACÍS FLORES³

RESUMEN:

Las más importantes redes sociales han respondido al contenido denunciado como *discurso de odio* dentro de las veinticuatro horas a partir de su publicación, un proceso apresurado que puede intercambiar una expresión tan controvertida por resultados rápidos, haciendo que el contenido de interés periodístico y la crítica terminen siendo eliminados. El impacto de la coerción extralegal es de gran alcance, puesto que las leyes nacionales están limitadas por las fronteras geográficas, a diferencia de la aplicación global de los términos de servicio. Las empresas pueden contrarrestarlo con claridad de definición, responsabilidad sólida, transparencia detallada y supervisión del defensor del pueblo.

¹ Artículo recibido el 24 de diciembre de 2019 y aprobado el 24 de enero de 2020.

² Abogado de los Tribunales y Juzgados de la República del Ecuador por la Universidad Central del Ecuador. Asesor adscrito del Comité Ejecutivo del Centro Educativo Latinoamericano de Investigaciones Sociales para el Siglo XXI. Secretario del Comité Ejecutivo del Centro Educativo Latinoamericano de Investigaciones Sociales para el Siglo XXI.
Correo electrónico: cesargalpaz@gmail.com
ORCID: 0000-0001-9349-7658

³ Egresado de la carrera de Jurisprudencia por la Pontificia Universidad Católica del Ecuador, sede Ibarra.
Correo electrónico: eddyvf_95@hotmail.com
ORCID: 0000-0002-2634-3605

PALABRAS CLAVE:

Libertad de expresión, censura, discurso de odio, redes sociales, términos de servicio.

ABSTRACT:

The most important social media platforms worldwide have replied to same content reported as “hate speech” through the next 24 hours since its posting date, an early procedure that might exchange a controversial expression for quick results, resulting in the fact that journalistic-interest content might be deleted. The impact in extra-legal pressure has reached a long range, due to national law is limited by geographic borders, unlike the global application if the terms of service. Companies can balance this issue with clear definitions, solid responsibility, detailed transparency and the vigilance of ombudsman.

KEYWORDS:

Freedom of expression, censorship, hate speech, social media terms of service

Introducción:

En 2008, el senador estadounidense JOSEPH LIEBERMAN se enfrentó a las compañías de Internet. La disputa se refería a la demanda del senador referida a que las plataformas eliminarán cientos de videos de capacitación de Al-Qaeda. El senador LIEBERMAN argumentó que al mantener los videos, las compañías de tecnología fueron cómplices en el reclutamiento de terroristas⁴. *YOUTUBE* de Google se mantuvo rápido en defensa del derecho de los usuarios al expresar puntos de vista impopulares. Como escribió JEFFREY ROSEN en ese momento, Nicole Wong de Google y sus colegas trabajaron *de manera impresionante para poner el compromiso a largo plazo de la compañía con la libre expresión por encima de sus intereses financieros a corto plazo*⁵.

Ignorar las demandas del senador era una estrategia segura: cualquier esfuerzo para proscribir contenido extremista probablemente fracasaría, dada la hostilidad de la Primera Enmienda frente a las regulaciones basadas en puntos de vista. La libertad de expresión estadounidense valora las decisiones de políticas guiadas en Silicon Valley mucho después del enfrentamiento con el senador LIEBERMAN. Las compañías de medios sociales miran rutinariamente a la doctrina de la Primera Enmienda al elaborar políticas en los términos de servicio referentes al discurso. TWITTER, un ejemplo de este espíritu, fue conocido como *el ala de la libertad de expresión del partido de la libertad de expresión*⁶.

⁴ LIEBERMAN, J. 2008.

⁵ LEE, T. 2008.

⁶ HALLIDAY, J. y WANG, T. 2008.

Desde el principio, el compromiso de las empresas de tecnología con la libre expresión admitió algunas excepciones. Los términos de servicio y las pautas de la comunidad prohibieron la pornografía infantil, el *spam*, el *phishing*, el fraude, la suplantación de identidad y las violaciones de derechos de autor⁷. Las amenazas, el acoso cibernético, la pornografía no consensual y el discurso de odio fueron prohibidos después de largas discusiones con grupos de defensa⁸. El objetivo era lograr un equilibrio adecuado entre la libertad de expresión y la prevención del abuso, al tiempo que se preservaba la participación del mercado de las plataformas.

Más recientemente, las compañías de medios sociales han revisado sus políticas de discurso, con respecto a la propaganda extremista y de odio. Sin embargo, a diferencia de los cambios anteriores, estas revisiones no fueron el resultado de las fuerzas del mercado, no se hicieron para satisfacer los deseos de los anunciantes y defensores. En su lugar, fueron adoptados para evitar la regulación europea amenazada. Después de los ataques terroristas en París y Bruselas, a finales de 2015, los reguladores europeos criticaron a las empresas de tecnología por no combatir el reclutamiento de terroristas en sus plataformas. Su mensaje era claro: las plataformas en línea enfrentarían penas civiles y penales onerosas, a menos que sus políticas y procesos resultaran en la rápida eliminación del discurso extremista⁹.

Las compañías tecnológicas se acomodaron a estas demandas, porque la regulación del discurso extremista era una posibilidad real. A diferencia de lo que ocurre en los Estados Unidos, en

⁷ TWITTER. 2016.

⁸ GUYNN, J. 2017.

⁹ GUYNN, J. 2017.

la Unión Europea no existe una fuerte presunción contra las restricciones de voz. El 31 de mayo de 2016, FACEBOOK, MICROSOFT, TWITTER y YOUTUBE firmaron un acuerdo con la Comisión Europea para eliminar el discurso de *odio* dentro de las veinticuatro horas, según corresponda, de acuerdo a sus términos del servicio¹⁰. Seis meses después, las mismas compañías anunciaron planes para una base de datos compartida de contenido extremista prohibido para su revisión y eliminación en otros lugares¹¹.

Casi una década después, los legisladores europeos lograron lo que el senador LIEBERMAN no pudo. Al insistir en los cambios en las reglas y prácticas de libre expresión de las plataformas, los reguladores de la UE han ejercido su voluntad en todo el mundo¹².

A diferencia de las leyes nacionales, que se aplican solo dentro de las fronteras de un país¹³, los términos de servicio se aplican donde se accede a las plataformas. De manera similar, mientras que los tribunales locales solo pueden ordenar plataformas para bloquear el material al que se accede en su jurisdicción, la base de datos de la industria tiene el potencial de resultar en una censura mundial.

¹⁰ UNIÓN EUROPEA. 2016.

¹¹ NEWTON, C. 2016.

¹² CORTE DE APELACIONES DE LOS ESTADOS UNIDOS. Caso Fairley vs. Andrews. 2009.

¹³ A diferencia de la UE, en los Estados Unidos, amenazar con regular el discurso protegido implica las protecciones de la Primera Enmienda. *Las penas amenazantes para el futuro del habla tienen el nombre de restricción previa, y una restricción previa es la primera violación de la primera enmienda.* CORTE DE APELACIONES DE LOS ESTADOS UNIDOS. Caso Fairley vs. Andrews. 2009.

Todo esto podría gozar de cierta justificación, si los reguladores de la UE centraran sus esfuerzos en el discurso proscrito en sus países. Pero éste no ha sido el caso. Los llamamientos para eliminar el discurso de odio se han disparado rápidamente para abarcar una expresión, que no viola la legislación europea existente, incluida la *radicalización en línea* y las *noticias falsas*¹⁴. Los funcionarios de la UE han insistido en un discurso de odio, que puede extenderse a la disidencia política y los acontecimientos de interés periodístico. En riesgo, la censura se arrastra a escala global.

Se ha explorado cómo los requisitos legales formales y la presión informal del gobierno pueden resultar en censura colateral: el silenciamiento de actores privados por parte de otros actores privados. El becario de libertad de expresión y fundador del Proyecto de la Sociedad de la Información de Yale, JACK BALKIN, advirtió:

*Actualmente, Internet se rige principalmente por los valores del régimen menos censorador: el de los Estados Unidos de América. Si los estados nacionales pueden imponer el filtrado, el bloqueo y la desvinculación globales, Internet se regirá por el régimen más censorador. Esto socavará el bien público global de una Internet gratuita*¹⁵.

El asalto al *bien público global de una Internet gratuita* ya está en marcha. Como muestra este artículo, la expresión digital se ajusta a la reglamentación sobre la libertad de expresión de la UE, con el discurso extremista y odioso como catalizadores.

¹⁴ McGOOGAN, C. 2016.

¹⁵ BALKIN, J. 2018.

Este artículo tiene tres partes. La Parte I expone la presión que enfrentan las empresas de tecnología para adaptar sus políticas de voz a las normas de la UE. Como muestra de la Parte I, analizaré el reciente retiro de Silicon Valley de un fuerte compromiso con la libertad de expresión, el cual tiene más que ver con la compulsión que con la elección. La Parte II explora las consecuencias, destacando el riesgo de que la censura se arrastre a escala global. La Parte III ofrece garantías diseñadas para contener la presión extralegal para el bien de la libre expresión.

1. El poder de la Unión Europea sobre las reglas de discurso privado

Después de una serie de ataques terroristas mortales y delitos de odio en el 2015, los legisladores europeos dijeron a las empresas de redes sociales que eran -en parte- responsables de la violencia. Desde su punto de vista, las plataformas en línea habían habilitado a extremistas violentos al darles acceso a posibles reclutas. Los legisladores europeos advirtieron a las compañías que enfrentarían multas penales y civiles, a menos que se eliminara el extremismo en línea.

Después del ataque a Charlie Hebdo, el presidente francés, François Hollande, pidió una legislación que hiciera a las plataformas de las redes sociales responsables del contenido *extremista* de los usuarios. El Ministro del Interior francés, Bernard Cazeneuve, siguió esa advertencia con reuniones en Silicon Valley. Las conversaciones con los ejecutivos de tecnología dieron algunos frutos: varias compañías acordaron continuar eliminando el contenido relacionado con el terrorismo. Esto fue solo el comienzo de las concesiones de Silicon Valley a los reguladores europeos.

a) Código de Conducta

El 3 de diciembre de 2015, la Comisión Europea estableció el Foro Europeo de Internet (en adelante el *Foro*). El objetivo era el desarrollo de un *enfoque conjunto y voluntario* para la detección y eliminación de la *incitación terrorista en línea y el discurso de odio*¹⁶. Entre los participantes, se encontraban funcionarios europeos, Europol y las empresas de tecnología FACEBOOK, Microsoft, TWITTER y YOUTUBE; el Comisario europeo de Migración, Asuntos del Interior y Ciudadanía comentó:

*Los terroristas están abusando de Internet para difundir su propaganda venenosa: eso debe detenerse. La asociación voluntaria que lanzamos hoy con la industria de Internet [tiene como objetivo] abordar este problema. Queremos resultados rápidos. Esta es una nueva forma de abordar este abuso extremista de Internet, y proporcionará la plataforma para compartir el conocimiento experto, [y] para desarrollar conclusiones rápidas y operativas [...]*¹⁷.

El Foro produjo resultados en poco tiempo. El 31 de mayo de 2016, la Comisión Europea anunció un acuerdo con las compañías tecnológicas titulado “Código de conducta para la lucha contra la incitación ilegal al odio en internet” (*acuerdo de odio o el Código*)¹⁸. Las compañías tecnológicas acordaron prohibir la *conducta de odio*, definida como un discurso que

¹⁶ *Ibíd.*

¹⁷ *Ibíd.*

¹⁸ CLARK, L. 2016.

incita a la violencia o al odio contra grupos protegidos¹⁹. Los informes de discursos de odio serían revisados dentro de las veinticuatro horas y eliminados si el discurso violaba los términos de servicio de las compañías. La Comisión Europea dejó en claro que realizaría revisiones periódicas del cumplimiento de las compañías tecnológicas con el acuerdo de discursos de odio. La Comisionada Europea de Justicia, Consumidores e Igualdad de Género, VERA JOUROVÁ, calificó el acuerdo de incitación al odio como algo esencial para combatir el uso de las redes sociales para *radicalizar a los jóvenes y difundir la violencia y el odio*²⁰.

En diciembre de 2016, la Comisión Europea emitió su primera evaluación del manejo de los informes de odio por parte de las empresas de tecnología y los comentarios no fueron positivos. Durante un período de seis semanas, doce organizaciones, que trabajaban en nombre de la Comisión, informaron sobre supuestos casos de incitación al odio y siguieron la respuesta de las compañías²¹. La Comisión Europea criticó la *tasa de éxito* de las compañías de tecnología, el número de solicitudes que resultaron en eliminación y puntualidad. Solo el cuarenta por ciento de las acusaciones se revisaron en veinticuatro horas y el veintiocho por ciento de los que se informaron como conducta odiosa, se eliminaron.

b) Base de Datos *Lista Negra*

Las autoridades de la UE han estado en contacto con compañías de redes sociales sobre el uso de sus servicios por parte de los

¹⁹ EUROPEAN COMMISSION. 2016.

²⁰ TOOR, A. 2016.

²¹ UNIÓN EUROPEA. 2016.

grupos terroristas durante los últimos siete años²². Durante años, el contacto se desarrolló de manera *ad hoc* con el cumplimiento de la ley, solicitando a las empresas que eliminen el contenido²³. El Reino Unido estableció una Unidad de Referencia de Internet contra el terrorismo (CTIRU, por sus siglas en inglés) para identificar e informar el *extremismo violento y no violento* a las plataformas en línea²⁴. De 2010 al 2015, CTIRU aseguró la eliminación de 249.091 piezas de contenido relacionado con el terrorismo²⁵.

Dado el éxito de los esfuerzos de CTIRU, Europol estableció su Unidad de referencia de Internet, que se describe como una *asociación [...] con el sector privado (para promover la 'autorregulación' de los proveedores de servicios en línea)*²⁶. Noventa y uno por ciento del contenido informado se ha eliminado²⁷. EUROPOL ha descrito el cumplimiento por parte del sector privado de sus solicitudes de eliminación como *voluntario*.

A fines de 2015, las compañías de medios sociales enfrentaron una creciente presión para sistematizar el proceso de eliminación. El contenido extremista debía eliminarse más rápido y de ser posible, antes de estar en línea. Una sugerencia fue la adopción de una base de datos de la industria, la cual permita la detección de imágenes, audio y archivos de video terroristas, violentos,

²² MURDOCK, J. 2016.

²³ BENESCH, S., y MCKINNON, R. 2012.

²⁴ CRAIG, S. y LLANSÓ, E. 2015.

²⁵ EU INTERNET REFERRAL UNIT. 2016.

²⁶ *Ibíd.*

²⁷ BAKER, J. 2016.

prohibidos²⁸. La base de datos recopilaría *hashes*, o huellas digitales únicas, de contenido prohibido para que pueda marcarse y eliminarse instantáneamente²⁹.

Inicialmente, Silicon Valley rechazó la idea. Los ejecutivos expresaron su preocupación de que el contenido valioso pudiera ingresar a la base de datos ya que no había un consenso claro sobre el significado de *discurso terrorista*. Otra preocupación era que los gobiernos podrían tratar de usar la base de datos para silenciar a sus críticos.

Seis meses después, las compañías tecnológicas cambiaron de rumbo y anunciaron planes para una base de datos de la industria para ayudar a prevenir la propagación de *imágenes terroristas violentas*³⁰. En un comunicado de prensa, las compañías tecnológicas señalaron que otras plataformas serían bienvenidas tan pronto como la base de datos esté en funcionamiento³¹. La Comisión Europea calificó la base de datos como el siguiente paso lógico en una asociación público-privada para combatir el extremismo³². En su opinión, la base de datos era *un importante paso adelante en la forma de una respuesta colaborativa de la industria para proteger a sus usuarios del contenido terrorista*³³. Google también señaló su esperanza de que la colaboración *conduzca a una mayor eficiencia, a medida que continuamos aplicando*

²⁸ WADDEL, K. 2016.

²⁹ NAKASHIMA, E. 2016.

³⁰ FACEBOOK. 2016.

³¹ MOODY, G. 2016.

³² EUROPEAN COMISSION. 2016.

³³ *Ibíd.*

*nuestra políticas para ayudar a frenar el acuciante problema mundial del contenido terrorista en línea*³⁴.

Sin embargo, las compañías tecnológicas sí emitieron algunas pautas. La base de datos detectaría contenido de *las imágenes y los videos terroristas más extremos y notorios [...], con mayor probabilidad de violar todas las políticas de contenido de nuestras respectivas compañías*. El material no se eliminaría de inmediato de los sitios de los participantes. En cambio, cada compañía revisaría el contenido incluido en la base de datos bajo sus propias políticas. Las compañías de tecnología reconocieron la importancia de directrices claras para garantizar que la base de datos de la industria no se utilice para censurar el discurso público³⁵.

c) Más coerción que elección

La normativa europea ha influido en la política en todo el mundo, a menudo reforzando la regulación global. Más obviamente, la UE proporciona estándares globales, cuando las regulaciones extranjeras imitan sus estrictas reglas. La influencia de la UE también se siente cuando las empresas adoptan sus estrictos estándares, como resultado de las fuerzas del mercado. En lo que se ha denominado el *efecto de Bruselas*, las empresas siguen las leyes de la UE en sus operaciones globales, porque les permite operar simultáneamente en los mercados internos más grandes del mundo y en jurisdicciones con estándares más laxos. En áreas como la privacidad de los datos y leyes antimonopolio, las empresas han ajustado sus operaciones globales a estándares

³⁴ SHACKFORD, S. 2016.

³⁵ CLARK, L. 2016.

Europeos más restrictivos para evitar la imposición de sanciones por cumplimiento fragmentario.

Sin embargo, las fuerzas del mercado no explican la reciente adopción, por parte de Silicon Valley, de las normas de discurso de la UE, relacionadas con el discurso extremista. Las compañías tecnológicas no cambiaron sus políticas y prácticas de libre expresión por razones de eficiencia. No intentaban atraer tarifas publicitarias, ni la aprobación de los defensores. Estaban respondiendo a las amenazas del gobierno: no haciendo cambios voluntarios en sus políticas.

Las empresas tecnológicas cedieron a las demandas de los reguladores europeos, porque sabían que sus amenazas no estaban inactivas. En la Unión Europea, a diferencia de los Estados Unidos, no existe una fuerte presunción contra las restricciones de la libre expresión³⁶. El discurso de odio y la propaganda extremista ya estaban prohibidos en algunos estados miembros de la UE³⁷.

Los proveedores de redes sociales sabían que si se aprobaran y aplicaran leyes onerosas, alojar contenido generado por el usuario sería prohibitivamente costoso. No importa con cuánta frecuencia los legisladores de la UE describan los cambios en plataformas digitales como decisiones *voluntarias*, solo pueden entenderse razonablemente como el producto de la coerción del gobierno.

La comparación de la dinámica de los cambios anteriores en las políticas de expresión de las empresas con los desarrollos recientes

³⁶ UNIÓN EUROPEA. 2010.

³⁷ BERSHIDSKY, L. 2016.

demuestra este punto. En entrevistas realizadas en 2010 y 2011, el personal de seguridad explicó que sus plataformas prohibían el discurso de odio porque se ajustaba al sentido de responsabilidad social corporativa de sus empleadores, a los deseos de los anunciantes y grupos de defensa³⁸. FACEBOOK, Xbox Live de MICROSOFT y otros anfitriones de contenido adoptaron diferentes definiciones del discurso de odio³⁹, desde el restrictivo (por ejemplo, el discurso dirigido a grupos marginados con violencia) hasta el amplio (por ejemplo, discurso que desprecia a los grupos marginados)⁴⁰. Las políticas corporativas del discurso de odio fueron un reflejo de las opciones comerciales, no de la presión del gobierno.

De manera similar, los cambios en las reglas de discurso privado en torno a la pornografía no consensual fueron un reflejo de las fuerzas del mercado. Ya en 2011, los grupos de defensa estaban presionando a las plataformas para cambiar sus reglas con respecto a la publicación de imágenes de desnudos sin el consentimiento de los sujetos⁴¹. Argumentaron que la pornografía no consensual no era buena para los negocios ni el resultado de la expresión sexual voluntaria.

El único aprovechamiento de los defensores fue su capacidad para unir a la opinión pública, lo que podría imponer costos modestos si algunos usuarios y anunciantes abandonaran sus plataformas. Después de que individuos destacados denunciaron la publicación no consensuada de sus imágenes de desnudos y de los medios

³⁸ CITRON, D. K., y NORTON. 2011.

³⁹ BURLACU, A. 2016.

⁴⁰ *Ibíd.*

⁴¹ CITRON, D. K. 2015.

de comunicación populares criticaron la falta de atención de las empresas al problema en el 2014, las principales plataformas en línea revisaron sus políticas para prohibir la práctica. Para estar seguros, los grupos de defensa tenían la capacidad de reunir el apoyo público detrás de la noción de que las empresas deberían prohibir la pornografía no consensuada. Pero, a diferencia de los gobiernos, no tenían ni el apalancamiento ni el poder para hacer que a las compañías les resulte prohibitivamente costoso ignorar su defensa. Por eso llevó una gran cantidad de tiempo y de manera crucial, presión del mercado, para que las plataformas cambien sus posiciones sobre la pornografía no consensuada.

Por el contrario, cuando las compañías tecnológicas cambiaron sus políticas y procedimientos de discurso con respecto al material odioso y extremista en el 2016, lo hicieron a la sombra de una regulación amenazada⁴². En lo que seguramente fue programado para disminuir el impacto de la revisión crítica de la Comisión Europea sobre el cumplimiento del acuerdo de odio, las compañías tecnológicas publicaron su anuncio sobre la base de datos de la industria el día antes de que la Comisión Europea publicara su informe. Como observó Eve Peyser, la base de datos de la industria fue sin duda el *esfuerzo rápido y fuerte que pidió la UE*⁴³. Los reguladores de la UE ejercieron su poder para imponer costos materiales en el discurso extremista para presionar la conformidad con sus normas de libre expresión.

Las demandas de los líderes europeos solo han aumentado desde el anuncio del acuerdo de odio y la base de datos de la industria. Después de una serie de ataques terroristas en Londres,

⁴² KAUFFMAN, G. 2016.

⁴³ SCOTT, M. 2016.

en el 2017, la Primera Ministra británica, Theresa May, y el presidente francés, Emmanuel Macron, amenazaron con multar a las compañías tecnológicas por *no eliminar la propaganda extremista de sus plataformas*⁴⁴. Pidieron a las plataformas de medios sociales que utilizaran la automatización para evitar la publicación, en lugar de confiar en los usuarios para marcar el contenido para su eliminación⁴⁵. Alemania cumplió sus amenazas al aprobar una ley que sanciona multas de hasta cincuenta y siete millones de euros si las empresas no eliminan el discurso de odio dentro de las veinticuatro horas posteriores a su informe⁴⁶.

En respuesta, Google anunció un plan de cuatro partes para abordar la propaganda terrorista que incluye un mayor uso de la tecnología para identificar videos relacionados con el terrorismo, contratar moderadores de contenido adicionales, eliminar publicidad en videos potencialmente objetables y dirigir a posibles reclutas terroristas hacia videos contra la radicalización⁴⁷. FACEBOOK anunció el uso de inteligencia artificial para detener la propagación de la propaganda terrorista y la contratación de 3000 personas más para revisar el discurso reportado como violaciones de los términos de servicio (TDS)⁴⁸. Estos esfuerzos parecen diseñados para evitar que los líderes europeos adopten o hagan cumplir leyes que castigarían a las empresas tecnológicas por no censurar adecuadamente el contenido extremista⁴⁹.

⁴⁴ PAULSON, A. y BOTKIN-KOWACKI, E. 2017.

⁴⁵ OFICINA DEL GABINETE y OFICINA DEL HOGAR. 2017.

⁴⁶ EDDY, M. y SCOTT, M. 2017.

⁴⁷ WALKER, K. 2017.

⁴⁸ BICKERT, M., y FISHMAN, B. 2017.

⁴⁹ CLARK, L. 2016.

2. La censura del contenido *Creep*

Sin lugar a dudas, la presión de la UE para eliminar conductas de odio y contenido terrorista violento podría ser beneficiosa. Los cambios en las prácticas de libre expresión de las empresas podrían evitar que los individuos marginados vean videos de Al-Qaeda, que piden la muerte de judíos y luego vayan a las sinagogas con armas. La base de datos compartida podría facilitar la rápida eliminación de videos espantosos de decapitación, evitando su propagación viral en todo el Internet. Con menos propaganda terrorista y menos discursos de odio en línea, podrían haber menos personas que se unan a combatientes de ISIS en Siria o que colocan bombas en mercados comerciales o en casas de culto.

Aunque estos beneficios potenciales no están en duda, también hay costos potenciales. Se podría interpretar que las políticas de TDS de las empresas prohíben el discurso mucho más allá del discurso que incita al odio (o que exhorta a la violencia) contra grupos vulnerables o contenido extremista violento. Podrían resultar en la eliminación global de los *tweets* de un funcionario del gobierno, podrían llevar a la eliminación mundial de sitios web que critiquen a los candidatos políticos⁵⁰, o podrían resultar en la suspensión global de los perfiles de FACEBOOK de los activistas de los derechos civiles⁵¹.

⁵⁰ WAMSLEY, L. 2017.

⁵¹ FACEBOOK prohibió temporalmente a un conocido activista y periodista de Black Lives Matter, Shaun King, que publicó mensajes racistas que recibió. El periodista LEVIN, SAM escribe: *FACEBOOK bloquea temporalmente al activista de Black Lives Matter después de publicar un correo electrónico racista*. LEVIN, S., en *The Guardian*. 2016.

Esta parte explica cómo esto podría ocurrir. Comienza por explorar el concepto de censura. Expone los costos del arrastre de la censura, incluida la supresión del debate legítimo y el discurso en contra que podría convencer a la gente de rechazar el fanatismo y la ideología terrorista.

a) Concepto

El término *creep* se refiere a la idea de que una herramienta diseñada para un propósito termina siendo utilizada para otro⁵². Las herramientas o programas diseñados para lograr un fin particular o para resolver un problema específico se extienden gradualmente a otros usos o contextos. En el contexto de la ingeniería, el fenómeno se denomina función *creep*⁵³. Dependiendo de las circunstancias, puede tener beneficios. Los nuevos usos de las tecnologías de propósito general pueden producir crecimiento económico, innovación y otras externalidades positivas⁵⁴; pero la función *creep* también puede ser costosa, teniendo efectos perniciosos que son tan graduales que eluden una revisión cuidadosa.

Puede considerarse dicha función involucrada en ciertos programas de vigilancia. Como FRANK PASQUALE y DANIELLE KEATS CITRON exploran en trabajos anteriores, los *fusion centers* comenzaron como asociaciones federales-estadales para combatir el terrorismo⁵⁵. Con el tiempo, su misión no se vio aminorada desde sus inicios antiterroristas para cubrir todos los

⁵² FRISCHMANN, B., y SELINGER, B. 2018.

⁵³ *Ibíd.*

⁵⁴ *Ibíd.*

⁵⁵ CITRON, D. K., y PASQUALE, F. 2011.

delitos, amenazas y riesgos⁵⁶. Aunque se obtuvo poca seguridad con la expansión del trabajo de los *fusion centers*, tuvo un costo, incluida la pérdida de la privacidad individual, el habla entrecortada y la redirección de recursos lejos de las misiones de inteligencia central⁵⁷.

Ahora paso al concepto de la censura, que se refiere a la expansión de las políticas del habla más allá de sus objetivos originales⁵⁸. Como explica PAUL BERNAL, *cuando se construye un sistema de censura para un propósito, se puede estar bastante seguro de que se utilizará para otros propósitos*⁵⁹. Esta sección explora el potencial preocupante de la censura adoptada bajo la presión de la UE. Explica los motivos de la censura y detalla sus costos para las libertades individuales y la aplicación de la Ley.

b) Causas

Varias tendencias fomentarán y exacerbarán la censura: ambigüedad de definición, cumplimiento global de las reglas de expresión de las compañías y opacidad de las prácticas de expresión privadas. Esta sección explora las implicaciones de estas tendencias.

1. Definición ambigua

La censura ocurre cuando las reglas del habla se basan en una terminología ambigua. Sin pautas claras y ejemplos específicos,

⁵⁶ *Ibíd.*

⁵⁷ *Ibíd.*

⁵⁸ PAULSON, A., y BOTKIN-KOWACKI, E. 2017.

⁵⁹ BERNAL, P. 2014.

los términos vagos son vulnerables a la revisión y expansión. Considere la definición del Código de *discurso de odio ilegal*: el discurso incita a la violencia o el odio contra un grupo o un miembro de dicho grupo basado en raza, religión, nacionalidad u origen étnico⁶⁰. Incitar al odio contra un grupo es un concepto ambiguo. Podría interpretarse para cubrir un discurso ampliamente entendido como odioso, como describir a los miembros de un grupo religioso como parásitos responsables del crimen y la enfermedad. Pero también podría entenderse que abarca un discurso que muchos caracterizarían como de interés periodístico. Dada la ambigüedad del término, la incitación al odio podría extenderse a las críticas de los católicos por encubrir la explotación sexual de niños por parte de los sacerdotes⁶¹. Podría interpretarse como una aplicación al discurso que desafía al fundamentalismo islámico por su homofobia o supresión de las mujeres⁶². Podría extenderse a discurso que expone el odio, el cual enfrentan las minorías raciales.

En el contexto del acuerdo sobre discurso de odio, la censura no es una posibilidad teórica. Ya está sucediendo. Los funcionarios europeos han combinado el *discurso de odio ilegal* con el contenido terrorista, discurso extremista e historias falsas. Al criticar la implementación del acuerdo por parte de las Compañías Tecnológicas, la Comisionada Europea de Justicia, Consumidores e Igualdad de Género señaló que las compañías no lograron eliminar *la radicalización en línea, el discurso ilegal de odio [y] las noticias falsas*⁶³. El debate legítimo podría caer

⁶⁰ COMISIÓN EUROPEA. 2016.

⁶¹ POST, R. 2009.

⁶² *Ibíd.*

⁶³ McGOOGAN, C. 2016.

fácilmente dentro de esa amplia caracterización del discurso de odio.

Que las definiciones vagas de discurso de odio se pueden usar para suprimir la disidencia legítima es una preocupación de hace tiempo. En particular, la definición de discurso de odio que figura en el acuerdo de discurso de odio ha estado sujeta a esta crítica durante mucho tiempo. Durante la redacción del Pacto Internacional de Derechos Civiles y Políticos en la década de 1940, los Estados Unidos de América (liderados por Eleanor Roosevelt) argumentaron que los dictadores podrían manipular la misma definición de discurso de odio para censurar la disidencia⁶⁴.

Como argumentó Roosevelt hace setenta años, los gobiernos pueden verse tentados a denunciar la disidencia política como discurso de odio con la esperanza de que las empresas lo eliminen. El defensor de los Derechos Humanos, JACOB MCHANGAMA, se ha preguntado si las compañías tecnológicas podrán resistir las *demandas inevitables* de los estados para eliminar el contenido que determinen como *odioso* o *extremista*⁶⁵. La resistencia puede ser difícil, especialmente si una parte significativa del mercado de las empresas proviene de los países que solicitan la eliminación de la palabra.

La base de datos de la industria plantea preocupaciones similares. La noción de contenido terrorista violento es vaga y, por lo tanto, sujeta a expansión. Aunque las compañías tecnológicas se han comprometido a incluir solo las *imágenes y videos terroristas*

⁶⁴ ASWAD, E. 2016.

⁶⁵ KAUFFMAN, G. 2016.

*más extremos y atroces*⁶⁶, sin embargo, es algo que no está claro. Las compañías tecnológicas actualmente tienen diferentes definiciones de contenido terrorista en sus términos de servicio, desde *contenido destinado a reclutar para organizaciones terroristas*⁶⁷ a *amenazas permanentes (directas o indirectas) a grupos de personas*⁶⁸.

Es probable que los gobiernos aprovechen la falta de claridad en el significado del contenido terrorista. Las compañías tecnológicas pueden enfrentar la presión de los actores estatales para incluir contenido gráfico y violento de todo tipo, no solo imágenes terroristas. En este sentido, un Ministro de Seguridad e Inmigración del Reino Unido ha argumentado que las plataformas deberían bloquear el contenido terrorista, incluso si no es ilegal porque la gente no quiere ver material *desagradable*⁶⁹.

Los partidarios de la base de datos de la industria han señalado el éxito de *Photo DNA* de MICROSOFT, que utiliza tecnología para eliminar la pornografía infantil de Internet⁷⁰. La base de datos de *Photo DNA*, administrada por el Centro Nacional para Niños Desaparecidos y Explotados (NCMEC, por sus siglas en inglés), no se ha transformado en una herramienta para la censura de otros tipos de material⁷¹. Pero la pornografía infantil es muy diferente de las imágenes terroristas. La pornografía infantil no es un concepto ambiguo, sus características se pueden definir e

⁶⁶ PÉREZ, S. 2016.

⁶⁷ YOUTUBE. 2019.

⁶⁸ TWITTER. 2019.

⁶⁹ CLARK, L. 2014.

⁷⁰ WADDELL, K. 2016.

⁷¹ KNIBBS, K. 2017.

identificar fácilmente. La base de datos de pornografía infantil es administrada por NCMEC, una organización con conocimiento experto y experiencia en pornografía infantil⁷². Como explicó Andy Sellars, la pornografía infantil *es realmente el único lugar donde los medios de comunicación son contrabando por su propia definición*⁷³.

En contraste, si el contenido equivale a material terrorista violento y atroz depende del contexto general, incluido el mensaje y la redacción precisa⁷⁴. Como explica Lee Rowland de la Unión Americana de Libertades Civiles (ACLU, por sus siglas en inglés), *los algoritmos no son buenos para determinar el contexto y el tono como el apoyo o la oposición, el sarcasmo o la parodia*⁷⁵. A diferencia de la clara ilegalidad de la pornografía infantil, el discurso terrorista violento puede ser precisamente eso, o por otro lado, noticias o defensa contra ideologías violentas⁷⁶.

⁷² JEONG, S. 2016.

⁷³ En abril de 2017, FACEBOOK anunció su adopción de nuevas herramientas para abordar el problema de la pornografía no consensual. FACEBOOK permitirá a los usuarios reportar fotos íntimas publicadas, sin el consentimiento de los *representantes especialmente capacitados* del equipo de operaciones de la comunidad del sitio, quienes revisarán las imágenes reportadas y las eliminarán, si violan los estándares comunitarios de FACEBOOK. HERN, A. 2017.

⁷⁴ WADDELL, K. 2016.

⁷⁵ *Ibíd.*

⁷⁶ *Ibíd.*

2. Delación global

Los cambios recientes en las políticas y prácticas privadas relacionadas a la regulación de la libre expresión corren el riesgo de convertir la delación en el valor predeterminado global. ¿Cómo es eso? Hay dos razones interrelacionadas: primero, la delación puede convertirse en la respuesta alternativa a los requerimientos de remoción de contenido censurable; y segundo, los acuerdos TDS a menudo se aplican globalmente.

Para ilustrar la probabilidad de que la eliminación se convierta en automática, tenga en cuenta el código de conducta del discurso de odio. La eliminación puede convertirse en la posición de respaldo en el discurso de odio para evitar las críticas y la nueva regulación. Una presunción de eliminación es ciertamente más económica que asumir los costos de las nuevas leyes. Por esta razón, el Secretario General del Consejo de Europa advirtió que las entidades privadas en las asociaciones público-privadas probablemente ejercerán un *control excesivo* sobre el contenido en línea para evitar la responsabilidad por la transmisión de contenido ilegal⁷⁷.

Una práctica predeterminada de eliminación puede ser reforzada por el requisito del acuerdo de discurso de odio de que las compañías tecnológicas respondan a los informes de discurso de odio dentro de las veinticuatro horas. La velocidad inevitablemente sacrifica la deliberación reflexiva. El corto plazo proporciona una razón adicional para que la eliminación se convierta en la respuesta predeterminada al discurso de odio informado⁷⁸.

⁷⁷ CONSEJO DE EUROPA. 2016.

⁷⁸ YORK, J. 2016.

La supuesta supresión de los discursos de odio por parte de las empresas está destinada a tener un impacto global porque los acuerdos TDS están involucrados en lugar de órdenes judiciales u otras formas de proceso legal. Los acuerdos de TDS suelen ser los mismos en todo el mundo⁷⁹. Por lo tanto, las decisiones de eliminar o bloquear el contenido como infracciones de TDS significan que el contenido se eliminará o se bloqueará en cualquier lugar donde se vea la plataforma⁸⁰.

Considerando el impacto diferente de una orden judicial y una violación de TDS, suponiendo que Francia procesó con éxito un cartel por incitar a la violencia contra los musulmanes, con una orden judicial en mano, las autoridades francesas podrían pedir a las plataformas en línea que bloqueen el discurso de odio ilegal. El acceso al contenido probablemente se bloquee en Francia y solo en Francia. Pero las violaciones de TDS proceden de una manera más sistemática. Los términos del servicio relacionados con el discurso de odio y extremismo se aplican a las operaciones globales de las compañías tecnológicas. Debido a que el acuerdo de discurso de odio está operacionalizado a través de los términos de servicio, una presunción para eliminación podría significar la eliminación del contenido web mundial.

La base de datos de la industria también tiene el potencial de incluir en la lista negra de contenido en todo el mundo. Las compañías pueden sentir la presión de eliminar las imágenes con hash que otros participantes hayan designado como expresiones extremistas que violan los términos del servicio⁸¹. Si ese fuera

⁷⁹ YOUTUBE. 2019.

⁸⁰ LLANSÓ, E. 2017.

⁸¹ *Ibíd.*

el caso, la base de datos de la industria podría convertirse en un programa de *borrado de todo*. La base de datos puede desaparecer o nunca aparecer en ningún lugar en línea.

Aunque el plan actual es que cada participante realice una revisión independiente del contenido marcado, puede cambiar a medida que las compañías tecnológicas y los futuros participantes confíen en las evaluaciones de cada uno. Dadas las posiciones dominantes en el mercado de los participantes actuales y la probabilidad que nuevas empresas sigan el liderazgo de cada uno, el contenido terrorista incluido en la base de datos sería invisible para la mayoría de los usuarios en línea. Si un contenido marcado se trata como una razón automática para bloquear el contenido en las principales plataformas, el contenido no existiría.

Esto probablemente sucederá cuando las empresas más pequeñas tengan acceso a la base de datos de la industria. Es probable que las empresas de nueva creación y otras compañías con limitaciones de capital carezcan de los recursos para realizar una revisión cada vez que se detecte contenido censurable en sus servidores⁸². Pueden inclinarse a eliminar el contenido incluido en la base de datos cuando aparezca en sus sitios.

Como advierte la Directora de Libertad de Expresión del Centro para la Democracia y la Tecnología, EMMA LLANSÓ, la base de datos de la industria *se convertirá en un objetivo para los gobiernos y actores privados que buscan suprimir el discurso en toda la web*⁸³. Esto le da a los gobiernos un punto único de presión que, si es exitoso, permitiría el bloqueo de todo tipo de

⁸² *Ibíd.*

⁸³ LLANSÓ, E. 2017.

contenido desfavorecido en Internet. SCOTT SHACKFORD señaló: *Una vez que una herramienta se puede usar para censurar, en masa, una foto violenta de algún terrorista del Estado Islámico, esa herramienta se puede usar para censurar cualquier cosa a grandes rasgos*⁸⁴.

3. Opacidad

Lo que agrava estas preocupaciones es la opacidad de las prácticas privadas del habla. El contenido odioso o terrorista se está eliminando fuera de los procesos gubernamentales formales. Cuando las autoridades de la UE presentan informes de discursos de odio bajo los términos del servicio, es a través del sistema de quejas de la empresa en lugar de un proceso administrativo o judicial. A diferencia de las solicitudes ante un órgano administrativo o judicial, no hay palancas oficiales para descubrir que un gobierno ha hecho la solicitud.

En los últimos años, las plataformas dominantes en línea han proporcionado informes de transparencia que permiten a los usuarios conocer la cantidad de solicitudes formales realizadas por países específicos para eliminar contenido. Pero esos informes de transparencia pueden no captar completamente las solicitudes gubernamentales para eliminar el discurso de odio o para agregar contenido extremista violento a la base de datos compartida de la industria. Debido a que a las empresas se les pide que eliminen el discurso odioso o extremista a través de sus procesos privados en lugar de los gubernamentales, no hay garantía de transparencia. Como se aborda en la Parte III, esas solicitudes pueden quedar

⁸⁴ SHACKFORD, S. 2016.

fuera de los esfuerzos voluntarios de transparencia emprendidos por compañías tecnológicas.

c. Riesgos

La censura del contenido *creep*, crea riesgos serios para la libertad de expresión global. La información puede eliminarse aunque sea esencial para un debate público significativo y para un informe completo de las noticias⁸⁵. Las personas necesitan hablar y escuchar a otros para gobernarse a sí mismos, eso incluye la capacidad participar libremente en todas las *formas de crear significado e influencia mutua*⁸⁶. Como escribió el comité editorial del Washington Post en respuesta a la eliminación de la propaganda terrorista de las empresas de medios sociales, *los ciudadanos de todos los países merecen saber qué está pasando en el mundo y lo que piensa la gente en ambos extremos del espectro, por difícil que sea para el estómago*⁸⁷.

Como ha afirmado el activista de Derechos Humanos, Aryeh Neier (quien huyó de los nazis con sus padres en 1939), *la libertad de expresión en sí misma es el mejor antídoto contra las doctrinas venenosas de quienes tratan de promover el odio*⁸⁸. El público debe poder ver u oír puntos de vista odiosos o extremistas para interrogar y contrarrestar esos puntos de vista. La expresión de ideas odiosas o extremistas permite a la sociedad afirmar fuertes normas sociales que las rechazan, por lo que aquéllos que se oponen a las ideologías de odio o terroristas pueden explicar y

⁸⁵ RADSCH, C. 2015.

⁸⁶ BALKIN, J. 2016.

⁸⁷ THE WASHINGTON POST. 2016.

⁸⁸ RADSCH, C. 2014, p. 85.

justificar sus objeciones. Incluso, si la mayoría de las personas que abrazan las ideas de odio pueden no estar abiertas al contra discurso, algunas sí lo pueden estar. Por ejemplo, las discusiones en línea fueron fundamentales para el rechazo de MEGAN PHELPS-ROPER a la odiosa ideología de su familia.

En 2009, MEGAN PHELPS-ROPER, parte de la Iglesia Bautista de Westboro, reunió una considerable cantidad de seguidores que tuitearon contra las personas parte del colectivo LGBT. Con el tiempo, se conectó en línea con personas que no estaban de acuerdo con ella. Algunos de sus interlocutores señalaron la crueldad de sus posiciones⁸⁹.

PHELPS-ROPER explicó que sus discusiones en TWITTER la llevaron a rechazar las odiosas opiniones de la Iglesia Bautista de Westboro. Como explicó en su charla en TED a sus amigos en TWITTER:

Se tomaron el tiempo para entender las doctrinas de Westboro y, al hacerlo, pudieron encontrar inconsistencias que extrañaba toda mi vida [...] La verdad es que la atención que me mostraron estos extraños en internet fue [...] evidencia de que las personas del otro lado no eran los demonios que me habían hecho creer⁹⁰.

Esas conversaciones en línea *alteraron la vida* para PHELPS-ROPER, porque la ayudaron a ver que los miembros de su iglesia no eran los *últimos árbitros de la verdad divina, sino seres*

⁸⁹ CHEN, A. 2015.

⁹⁰ PHELPS-ROPER, M. 2017.

*humanos defectuosos*⁹¹. El trabajo de su vida es ahora hablar en contra del fanatismo⁹².

Aunque los riesgos en contra de la libertad de expresión son reales, hay otros costos más allá de este ámbito. La eliminación del discurso extremista puede dificultar que la policía haga su trabajo. Las investigaciones sobre terrorismo a menudo se basan en pistas dejadas en la actividad de las redes sociales. Por lo tanto, puede ser difícil investigar un posible acto de terrorismo si se eliminan inmediatamente las pruebas en línea.

Protección contra la censura de contenido *creep*

El acuerdo de discurso de odio y la base de datos de la industria están aquí para quedarse: su desmantelamiento a corto plazo es muy poco probable. Es probable que la UE continúe exigiendo más cambios *voluntarios* para forzar la conformidad con las normas de expresión deseadas. Para los posibles censores estatales, las *asociaciones público-privadas* son cursos de acción fructíferos. Aseguran la adopción de preferencias gubernamentales, sin la carga del proceso formal. Los actores estatales disfrutan de la ventaja del poder gubernamental, mientras evitan el desorden de los debates políticos y las audiencias judiciales. Los procesos de informes privados también permiten una censura mucho mayor de lo que permitiría la ley. Las definiciones de plataformas de discurso de odio se pueden ampliar para cubrir todas las formas de contenido desfavorable, incluida la disidencia política. Los reguladores de la UE no tienen ninguna razón para apartarse de este enfoque, dado el éxito de los esfuerzos extralegales.

⁹¹ *Ibíd.*

⁹² *Ibíd.*

Las leyes diseñadas para verificar el potencial de la censura son poco probables. REBECCA MCKINNON lo expresó bien:

La colusión corporativa en la vigilancia y censura del gobierno es poco probable que se resuelva con la aprobación y la aplicación de las leyes, incluso con los gobiernos más bien intencionados y democráticos⁹³.

En última instancia, Silicon Valley puede ser la mejor protección contra la censura.

Como se describe en esta Parte, las compañías tecnológicas deben adoptar políticas y procedimientos especiales para protegerse contra la extralimitación gubernamental. El objetivo sería mejorar la claridad, la responsabilidad y la transparencia de los esfuerzos de censura. Estas propuestas se ofrecen con la esperanza de que los ejecutivos las vean como una forma efectiva de contener la presión de censura de la UE y transmitir su apoyo a la expresión de los usuarios.

a) Definiciones

La claridad en la definición, el significado y la aplicación de los términos *discurso de odio* y *material terrorista* ayudaría a contener el avance de la censura. En 2011, DANIELLE CITRON y HELEN NORTON argumentaron que cuando las compañías prohíben el discurso de odio, deben explicar con precisión qué significa el término y por qué se ha prohibido⁹⁴. En esta opinión, los usuarios necesitaban esta información para comprender sus

⁹³ BENESCH, S. y MCKINNON, R. 2012.

⁹⁴ CITRON, D. K., y NORTON, H. 2011.

derechos y responsabilidades al usar plataformas. La claridad definitoria sirve a otro objetivo: evitar que las prohibiciones privadas de discurso de odio sean apalancadas para silenciar la expresión legítima.

CITRON y NORTON establecieron posibles definiciones para el discurso de odio basado en los principios de la ley de derechos civiles y de responsabilidad civil de los Estados Unidos⁹⁵. Por ejemplo, el discurso de odio podría definirse como aquel que amenaza o fomenta la violencia contra grupos tradicionalmente subordinados (o miembros específicos de grupos)⁹⁶. Podría definirse como un discurso que intencionalmente inflige angustia emocional, expresión que está dirigida individualmente, especialmente amenazante o humillante, repetida y dependiente de material sensible o escandaloso⁹⁷. Podría definirse como un discurso que se elevaría al nivel de acoso procesable en el lugar de trabajo según las leyes de derechos civiles⁹⁸. Sin embargo, esas definiciones pueden ser demasiado limitadas para satisfacer a los reguladores de la UE.

Las empresas tecnológicas podrían buscar orientación en el Derecho internacional de los Derechos Humanos⁹⁹. Sin embargo, el problema de la ambigüedad persistiría, porque el Derecho de los Derechos Humanos contiene normas excepcionalmente flexibles¹⁰⁰. Otra posibilidad es consultar el trabajo del Secretario

⁹⁵ *Ibíd.*

⁹⁶ *Ibíd.*

⁹⁷ *Ibíd.*

⁹⁸ *Ibíd.*

⁹⁹ CRAIG, S., y LLANSÓ, E. 2015. .

¹⁰⁰ Explicando que el Derecho internacional no abarca una definición

General del Consejo de Europa, quien está ideando *estándares europeos comunes (para el discurso de odio y el material terrorista) para proteger mejor la libertad de expresión en línea*¹⁰¹. Es de esperar que esos esfuerzos proporcionen definiciones, que reduzcan la maleabilidad de esos términos.

Los grupos de Derechos Humanos han expresado serias reservas sobre el acuerdo de discurso de odio, porque el discurso de odio no está bien definido. La Comisionada de Justicia de la UE, VERA JOUROVA, respondió a sus objeciones, señalando la Decisión Marco de la UE, la jurisprudencia del Tribunal Europeo de Derechos Humanos y la Ley de los Estados parte¹⁰². Como mínimo, las empresas deberían mantener a la Comisión Europea en esa posición y mirar a aquellas fuentes en la elaboración de sus propias definiciones de discurso de odio y material extremista.

¿Qué pasa con la base de datos de la industria de contenido terrorista prohibido? Las leyes de los Estados parte de la UE brindan poca orientación porque son excepcionalmente

universalmente aceptada de odio, mientras que los Estados parte generalmente definen el discurso de odio como expresiones que expresan odio o antipatía hacia otros grupos sociales, un estándar que dificulta el trazado de líneas, si no imposible. Como ABRAMS, F. explica en su libro más reciente *El alma de la primera enmienda*, el Tribunal Europeo de Derechos Humanos ha confirmado las condenas de discurso de odio, que implican críticas a los políticos y opiniones intolerantes expresadas por ellos. RADSCH, C. 2014.

¹⁰¹ CONSEJO DE EUROPA. 2016.

¹⁰² La Comisionada JOUROVA no se ha adherido a esta definición al hablar sobre el acuerdo de discurso de odio. Como se explicó en la Parte I, ella ha sugerido que las noticias falsas, la radicalización en línea y la propaganda terrorista deberían eliminarse, de conformidad con el acuerdo. JOUROVA, V. 2016.

amplias¹⁰³. Las empresas tecnológicas han dicho que solo tienen la intención de denunciar las *imágenes y videos terroristas más extremos y atroces -contenido con mayor probabilidad de violar todas nuestras políticas de contenido-*¹⁰⁴. Las inquietudes sobre el avance de la censura pueden abordarse con políticas que expliquen exactamente lo que significa la frase *las imágenes y videos terroristas más extremos y atroces*. Las políticas deberían proporcionar ejemplos específicos de contenido, que merezca esa designación. Dicha claridad, explicaciones y ejemplos pueden ayudar a prevenir la ampliación gradual de los estándares, que rigen el contenido que se incluye en la base de datos.

A medida que las empresas tecnológicas trabajan en sus definiciones de discurso de odio y material extremista, podrían considerar la inclusión de grupos de Derechos Humanos y académicos en sus esfuerzos¹⁰⁵. Basándose en el trabajo del Cyberhate Working Group¹⁰⁶ y Global Network Initiative, una iniciativa de trabajo conjunto para ayudar a las empresas a articular lo que constituye el discurso de odio o el material terrorista. Los grupos de libertades civiles han defendido un papel para ayudar a las empresas a comprender *los diversos significados dados al*

¹⁰³ Las leyes nacionales difieren en cuanto a qué contenido terrorista se considera ilegal y en qué contextos; algunas de esas normas son tan amplias, que se han utilizado para encarcelar a periodistas, blogueros y defensores de los Derechos Humanos.

¹⁰⁴ CLARK, L. 2016.

¹⁰⁵ Criticando la presencia de EUROPOL en Internet y otras autoridades que presionan a las empresas para eliminar el contenido terrorista o extremista y, por lo tanto, permiten a los estados regular y reprimir el discurso en línea indirectamente. EUROPEAN DIGITAL RIGHTS. 2016.

¹⁰⁶ KLONICK, K. 2014.

*‘extremismo violento’ y conceptos relacionados, y el impacto potencial de la ambigüedad en esta área en la promoción y protección de los derechos humanos*¹⁰⁷.

Esas definiciones, diseñadas para moderadores de contenido, deben compartirse públicamente para que los gobiernos puedan comprender los límites de los esfuerzos para eliminar el discurso, bajo los acuerdos de TDS. Con este conocimiento, los gobiernos podrían estar menos envalentonados para presionar a las empresas a ampliar las políticas de discurso de odio o terrorismo más allá del reconocimiento. Algunos, incluso, podrían reconsiderar el intento de silenciar la expresión impopular, pero protegida.

b) Responsabilidad sólida

Las solicitudes de remoción hechas por las autoridades estatales (u organizaciones no gubernamentales que actúen en nombre del estado) deben estar sujetas a una revisión rigurosa. Para empezar, se debe exigir a los funcionarios del gobierno que se identifiquen, cuando denuncien contenido por violaciones de TS. Las plataformas en línea deben saber que están tratando con autoridades gubernamentales o sus sustitutos. Debe haber un canal de informes separado para las autoridades gubernamentales y cualquier organización que trabaje en nombre de un estado. Por ejemplo, TWITTER tiene *canales de admisión (dedicados para la aplicación de la ley y otros reportes autorizados* para presentar *solicitudes legales*¹⁰⁸. Ese es un buen comienzo. Todas

¹⁰⁷ ORGANIZACIÓN DE LAS NACIONES UNIDAS. Comentarios escritos del Artículo 19 sobre el Informe del Relator Especial sobre la Promoción y Protección del Derecho a la Libertad de Opinión y Expresión, 2015.

¹⁰⁸ El canal de informes por separado se aplica a las *solicitudes legales*,

las solicitudes de eliminación (*solicitudes legales* e informes de TDS) deben proceder a través de ese canal.

¿Qué pasa con el proceso de revisión en sí? Como KATE KLONICK ha documentado¹⁰⁹, FACEBOOK tiene procesos de revisión internos extensos con instrucciones detalladas sobre qué contenido está prohibido y bajo qué circunstancias¹¹⁰. Los empleados internos o compañías externas como Sutherland y Deloitte toman decisiones sobre la moderación de contenido, en nombre de la empresa¹¹¹.

Para garantizar que los términos de servicio se cumplan de manera uniforme, los moderadores reciben una amplia capacitación sobre las reglas, así como sobre posibles prejuicios culturales y reacciones emocionales¹¹².

Otras plataformas en línea abordan la moderación de contenido de una manera igualmente cuidadosa. Como parte de ese proceso, todas las compañías deben asegurarse de que las solicitudes gubernamentales se vean a través de una lente especial. Cuando las autoridades gubernamentales buscan suprimir el discurso bajo los términos de servicio, los moderadores de contenido deben

que incluye órdenes judiciales y solicitudes formales de agencias gubernamentales y agentes del orden. TWITTER. 2017.

¹⁰⁹ KLONICK, K. *op. cit.*, pp. 43-57.

¹¹⁰ En 2012, se filtró en línea una versión del manual de instrucciones de FACEBOOK para revisores. Como muestra KLONICK, K., desde entonces FACEBOOK ha desarrollado instrucciones aún más amplias y detalladas para las personas que revisan las quejas de abuso. Véase CITRON, D. K. 2012.

¹¹¹ KLONICK, K. *op. cit.*, p. 49.

¹¹² *Ibíd.* pp. 51-53.

revisar las solicitudes con una presunción contra la eliminación, o al menos una buena dosis de escepticismo¹¹³. Los moderadores de contenido deben recibir capacitación sobre la censura, incluidos los esfuerzos gubernamentales pasados y presentes para silenciar a los críticos. La capacitación debe centrarse en cómo distinguir el discurso de odio prohibido o el material terrorista del contenido de interés periodístico. Esta no es una tarea fácil, pero crucial.

Las decisiones relacionadas con las solicitudes del gobierno deben ir acompañadas de una explicación: si los funcionarios encargados de tomar decisiones tienen que articular sus razones probablemente piensen más detenidamente sobre sus decisiones¹¹⁴. Cuando un moderador decide otorgar una solicitud del gobierno para la eliminación, basada en una violación de TDS, esa decisión debe pasar automáticamente a través de una segunda capa de revisión. Las personas cuyo discurso se elimina deben ser notificadas sobre la eliminación y se les debe dar la oportunidad de apelar¹¹⁵.

El requisito del acuerdo sobre el discurso de odio de que las compañías tecnológicas resuelvan los informes dentro de las veinticuatro horas hará que sea difícil y costoso implementar capas adicionales de revisión. Con moderadores bien entrenados,

¹¹³ ORGANIZACIÓN DE LAS NACIONES UNIDAS. Comentarios escritos del Artículo 19 sobre el Informe del Relator Especial sobre la Promoción y Protección del Derecho a la Libertad de Opinión y Expresión, 2016.

¹¹⁴ CITRON, D. K. *op. cit.*, 2008.

¹¹⁵ Las plataformas en línea dominantes hacen justamente eso para las decisiones que resultan en la eliminación del discurso. Muchos proporcionan un proceso similar a las apelaciones para que los usuarios puedan presentar sus objeciones.

políticas detalladas y una presunción contra la eliminación, la puntualidad puede ser un problema menor.

Las decisiones reales, por supuesto, pueden provocar la ira de la Comisión Europea, puesto que no estará satisfecha si no se otorgan las solicitudes de eliminación, tal como lo sugirió la revisión reciente. Esa tensión se mantendrá hasta que las empresas tecnológicas y los usuarios europeos puedan transmitir a los comisionados y legisladores sus preocupaciones sobre la libre expresión, tal como lo hizo Apple en el debate sobre cifrado en los Estados Unidos de América. Si los grupos de Derechos Humanos tuvieran la capacidad de actuar como asesores y vigilantes, entonces podrían ayudar a explicar al público los riesgos reales de la censura.

¿Qué pasa con la base de datos de la industria de contenido extremista, violento, prohibido? Como en la Parte anterior se exploró, la base de datos de la industria podría convertirse en una lista negra. Una vez incluido en la base de datos, es posible que el contenido nunca termine apareciendo en alguna de las principales plataformas. La inclusión en la base de datos tiene un mayor potencial para silenciar sistemáticamente un discurso disidente de las solicitudes individuales para eliminar contenido en términos de servicio. Como resultado, las protecciones fuertes son esenciales para evitar que los gobiernos coopten la base de datos.

Una opción es que las empresas adopten una regla general de que los gobiernos no pueden contribuir con etiquetas para la inclusión de contenido a la base de datos. Como argumenta EMMA LLANSÓ, las empresas tecnológicas deberían establecer que *declaren de manera clara e inequívoca, que bajo ninguna circunstancia aceptarán una contribución de una etiqueta de*

*alguien que actúe en nombre de un gobierno*¹¹⁶. Para evitar los esfuerzos para socavar esa prohibición, podrían anunciar cualquier esfuerzo o solicitud de un gobierno para socavar esa política¹¹⁷. Otra alternativa es que las empresas sometan una solicitud del gobierno a varias capas de revisión y condicionen la presentación a la aprobación del personal superior¹¹⁸. La responsabilidad rigurosa puede ayudar a prevenir la censura sistematizadora del discurso legítimo a instancias de los gobiernos europeos y otras naciones.

c) Transparencia significativa

Otro control de la censura es que las empresas presenten informes detallados sobre los esfuerzos gubernamentales para censurar el discurso de odio y el material extremista a través de medidas informales. Los informes de transparencia permiten la conversación pública sobre la censura. A su vez, los usuarios europeos pueden contactar a los legisladores con inquietudes sobre los intentos de usar compañías tecnológicas como representantes de censura. Cuanto más se enteren los usuarios sobre los esfuerzos de las empresas para proteger sus libertades fundamentales, más confianza tendrán los usuarios en los proveedores en línea e incluso, los defensores de los derechos humanos pueden llamar la atención sobre las preocupaciones sobre la censura. Finalmente, los informes de transparencia pueden generar *una discusión productiva sobre el uso apropiado y los límites de la autoridad (estatal)*¹¹⁹.

¹¹⁶ LLANSÓ, E. 2016.

¹¹⁷ *Ibíd.*

¹¹⁸ *Ibíd.*

¹¹⁹ LOVITT, L. 2016.

La Electronic Frontier Foundation comenzó a rastrear hasta qué punto las empresas eran transparentes sobre la frecuencia con la que bloquean o eliminan contenido o cuentas en el 2015. En ese momento, Google era la única compañía tecnológica, que proporcionaba informes de transparencia. Desde entonces, muchas más plataformas se han esforzado por proporcionar cierta transparencia sobre las solicitudes del gobierno para suprimir el discurso. Revisar las características exitosas de los esfuerzos actuales puede ayudar a formular un camino a seguir.

TWITTER ha sido aclamado por sus esfuerzos de transparencia y con razón. El Informe de Transparencia del 2016 de la compañía detalla el número de solicitudes legales para la eliminación de contenido según el país¹²⁰. De manera crucial y única, revela el número de solicitudes gubernamentales que buscan la eliminación del contenido de terrorismo por violaciones de TDS¹²¹. TWITTER está *trabajando activamente para expandir* sus informes para los gobiernos sobre todas las solicitudes de TDS *no legales conocidas [...] a través de su estándar [...] canales de admisión, tales como solicitudes para eliminar cuentas de suplantación y otro contenido, que viola las reglas contra el abuso*¹²². Aunque el Informe de Transparencia 2016 de FACEBOOK proporcionó información importante a los usuarios, no incluyó la cantidad de solicitudes gubernamentales que buscaban la eliminación de contenido, basado en sus términos de servicio y estándares comunitarios¹²³.

¹²⁰ TWITTER, 2016.

¹²¹ TWITTER recibió 716 informes de gobiernos relacionados con 5929 cuentas y que el 85% fueron eliminados por violaciones de TDS relacionadas con el extremismo violento). *Ibid.*

¹²² *Ibid.*

¹²³ Pero el Informe de Transparencia 2016 de FACEBOOK proporcionó de

Al igual que TWITTER ha hecho para el contenido terrorista y espera hacer mucho más en el futuro, los informes de transparencia corporativa deberían detallar el número, el tema y los resultados de todas las solicitudes del gobierno para eliminar contenido por violaciones de TDS¹²⁴. Si se permite a los gobiernos solicitar la adición de etiquetas a la base de datos de la industria, los informes de transparencia deben incluir detalles sobre esas solicitudes. Aunque la transparencia no puede resolver el problema de la censura, puede ayudar a contenerla, especialmente si se adoptan estándares sólidos y procedimientos de responsabilidad sólidos.

d) Defensores del Pueblo

Un problema agudo relacionado con la censura, es su potencial para suprimir contenido de interés periodístico. Las solicitudes de eliminación del gobierno pueden tratar de eliminar el terrorismo o el contenido de odio cuya publicación sea de interés público. Para abordar esta inquietud, las empresas deberían considerar contratar o consultar a consejeros externos, defensores de derechos humanos cuyo trabajo principal es el proceso de recopilación de noticias. Dichas personas, que también son conocidos como editores públicos, trabajan para *proteger la libertad de prensa* y promover el *periodismo de alta calidad*, el papel es *actuar [...] en el mejor interés de los consumidores de noticias*.

manera útil ejemplos de solicitudes gubernamentales para eliminar contenido según las leyes locales y su respuesta. Por ejemplo, Francia le pidió a FACEBOOK que elimine una fotografía que muestra a las víctimas del ataque terrorista en la sala de conciertos Bataclan. Como explicaba el informe, la foto no violaba los estándares de la comunidad de FACEBOOK, pero debido a que violaba la ley francesa, el acceso a la foto estaba bloqueado en Francia. FACEBOOK. 2016.

¹²⁴ FREEDOM ONLINE COALITION. 2016.

Los defensores del pueblo deben tener un papel especial en la evaluación de las solicitudes de eliminación del gobierno hechas a través de canales informales, tales como los términos de servicio o la base de datos de la industria. Pueden ayudar a identificar solicitudes, las cuales eliminarían material, que es importante para el debate público y el conocimiento. Entonces también, debido a que la base de datos de la industria plantea preocupaciones especiales sobre la supresión de la expresión, el defensor del pueblo podría revisar todas las contribuciones a la base de datos con el interés público en mente¹²⁵.

e) TDS adaptados geográficamente

¿Podrían las compañías tecnológicas limitar el impacto de la presión de la UE al negarse a incorporar las normas de la UE en los términos de servicio? Las empresas podrían adaptar los términos de servicio y las directrices de la comunidad a países y regiones específicos. Esto mapearía las normas de expresión en países y regiones, en lugar de imponer un modelo único para todo el mundo.

Las políticas que balcanizan las prácticas en línea corren el riesgo de poner en peligro el libre flujo de información. Considere el impacto de las leyes que requieren que las compañías almacenen los datos recopilados de los ciudadanos de un país dentro de sus fronteras. Las leyes de localización de datos, como se las conoce, ponen todas las interacciones en línea, bajo el control de los gobiernos locales, lo que es especialmente preocupante en el caso de los regímenes autoritarios.

¹²⁵ La base de datos debe ser accesible y analizada por expertos independientes, familiarizados con la libre expresión y la política.

Adaptar los términos de servicio por país o región corre el riesgo de restringir la libre expresión. Pero también tendría una ventaja importante, evitaría la eliminación o el bloqueo de la libre expresión en países con mayor protección, mientras que permite la eliminación o bloqueo en países con normas de habla más restrictivas.

CONCLUSIÓN

Silicon Valley tiene una larga historia de adoptar normas de expresión de estilo estadounidense. Durante más de una década, cuando las plataformas de redes sociales admitieron excepciones a la noción de que la información debería ser gratuita, tuvieron cuidado de asegurarse de que esas excepciones se mantuvieran limitadas. Esto era cierto si el problema era la suplantación, el acoso o la pornografía no consensuada. La libre expresión se ha incrustado durante mucho tiempo en la cultura corporativa de las empresas tecnológicas. A raíz de los recientes ataques terroristas y el resurgimiento de los grupos de odio, los reguladores europeos han presionado a las compañías tecnológicas para que cambien sus reglas y prácticas de discurso para eliminar el material extremista, que aparece en sus plataformas. Algunos de los cambios resultantes pueden hacer algo bueno; después de todo, una base de datos de la industria que marca la eliminación de videos que contienen *listas de asesinatos* o instrucciones de bombas puede prevenir la violencia. El acuerdo de discurso de odio podría conducir a la eliminación rápida de publicaciones, que exigen la muerte de minorías religiosas, lo cual si se permite, podría inspirar violencia física en los miembros de esos grupos

Pero los cambios en las normas y prácticas privadas deben entenderse y evaluarse, según lo que son: de conformidad

obligada con las normas del habla europea, a pesar de que las protestas de los reguladores de la UE no son voluntarias, ni el producto de asociaciones significativas público-privadas. En cambio, son el resultado de la coerción gubernamental, que ocurre fuera del Estado de Derecho. La presión sobre los estados ahora es diferente, lo cual ha provocado cambios que corren el riesgo de que la censura mundial se arrastre. Debido a que los gobiernos están utilizando términos de servicio para lograr sus fines, la supresión resultante del discurso será global.

Este artículo ofrece posibles salvaguardas para evitar la censura. Las empresas pueden y deben adoptar protecciones profilácticas diseñadas para manejar la presión extralegal por el bien de la libre expresión. Una mayor claridad, responsabilidad, transparencia y supervisión ayudarían a verificar los esfuerzos de la UE para censurar el discurso a escala global. A medida que las empresas evalúan estas sugerencias, un enfoque de múltiples partes interesadas podría proporcionar experiencia crucial a las empresas tecnológicas, a medida que desarrollan procesos para resistir el avance de la censura al tiempo que combate los peligros reales del discurso de odio y el material extremista. Es probable que exista una solución concebida e impuesta principalmente por alguna combinación de gobierno y corporación ser inadecuado para abordar los complejos problemas planteados por el problema del uso extremista de las redes sociales. Debe incluir expertos en extremismo y odio, víctimas de violencia extremista, estudiosos de la disidencia política y organizaciones de derechos civiles y libertades civiles que profundizarían los esfuerzos de empresas, como FACEBOOK, que está aportando experiencia contra el terrorismo internamente¹²⁶ y TWITTER, cuyo Consejo

¹²⁶ MILLS, C. 2016.

de Confianza y Seguridad brinda asesoramiento sobre posibles propuestas sobre abuso en línea¹²⁷.

Como lo ilustra la lucha de Apple con el gobierno de EE.UU. por el cifrado y como se reafirmó el apoyo unánime de Silicon Valley para esa posición, las empresas de tecnología disfrutaron del apoyo público, cuando defienden las libertades fundamentales. Las sugerencias esbozadas en este artículo, por lo tanto, pueden ser recibidas positivamente.

¹²⁷ TWITTER. *op. cit.*, 2018.

REFERENCIAS BIBLIOGRÁFICAS

ASWAD, EVELYN. “El papel de las empresas de tecnología de los EE. UU. Como el nuevo implante de discursos sobre el odio de Internet en Europa”, *Columbia Human Rights Law Review*. N.º 1, Año 6, 2016.

BAKER, JENNIFER. “Europol’s Online Censorship Unit Is Hazard and Unaccountable, says NGO”. *Ars Technica*. Disponible e la siguiente página web: <https://arstechnica.com/tech-policy/2016/07/europol-iru-extremist-content-censorship-policing/>. (última consulta, 4 de julio de 2019)

BALKIN, JACK M. “Libertad de expresión en la Sociedad Algorítmica. Big Data, gobernanza privada y nueva regulación del discurso escolar”. *Law Review*. Davis School of Law, U. of C., N.º 51, 2018.

_____. “The Cultural Democracy & First Amendment”. *North Western Univesity Law Review*. N.º 110, 2016.

BENESCH, SUSAN, y MCKINNON, REBECCA. “The Innocence of YouTube”. *Foreign Policy Review*. Disponible en la siguiente página web: <http://foreignpolicy.com/2012/10/05/the-innocence-of-youtube/>. (última consulta, 5 de octubre de 2019)

BERNAL, PAUL. “Censorship and Surveillance”, *Paul Bernal’s Blog*. Disponible en la siguiente página web: <https://paulbernal.wordpress.com/2014/09/25/censorship-and-surveillance/>. (última consulta, 25 de septiembre de 2019)

BERSHIDSKY, LEONID. “Europe Gets U.S. Tech Leaders to Self-Censor”, *Bloomberg View*. Disponible en la siguiente página web: <https://www.bloomberg.com/view/articles/2016-05-31/europe-gets-u-s-tech-leaders-to-self-censor>. (última consulta, 31 de mayo de 2019)

BICKERT, MONIKA, y FISHMAN, BRIAN. “Preguntas difíciles: cómo contrarrestar el terrorismo”, *Facebook Newsroom*. Disponible en la siguiente página web: <https://newsroom.fb.com/news/2017/06/how-wecounter-terrorism/>. (última consulta, 15 de junio de 2019)

BURLACU, ALEXANDRA. “Microsoft Tackles Online Hate Speech With New Tools And Resources To Combat Abuse”, *TechTimes*. Disponible en la siguiente página web: <http://www.techtimes.com/articles/175274/20160827/microsoft-tackles-online-hate-speech-with-new-tools-and-resources-to-combat-abuse.htm>. (última consulta, 27 de agosto de 2019)

CHEN, ADRIAN. “Dejar de seguir”, *New Yorker Magazine*. Disponible en la siguiente página web: <http://www.newyorker.com/magazine/2015/11/23/conversion-via-twitter-westboro-baptist-church-megan-phelpsroper>. (última consulta, 23 de noviembre de 2019)

CITRON, DANIELLE KEATS. “Online Engagement on Equal Terms”, *Boston University Law Review*. Disponible en la siguiente página web: <https://www.bu.edu/bulawreview/citron-online-engagement-on-equal-terms/>. (última consulta, 19 de octubre de 2019)

_____. “Technological Due Process”, *Washington University Law Review*. N.º 85, 2008.

_____. y NORTON, HELEN. “Intermediarios y discurso del odio. Fomentando la ciudadanía digital para nuestra información”, *Boston University Law Review*. N.º 91, 2011.

_____. y PASQUALE, FRANK. “Responsabilidad de la red para el aparato de inteligencia doméstica”, *Hastings Law Journal*. N.º 62, 2011.

CLARK, LIAT. “Facebook y Twitter deben abordar el discurso de odio o enfrentar nuevas leyes”, *Wired UK*. Disponible en la siguiente página web: <http://www.wired.co.uk/article/us-tech-giants-must-tackle-hate-speech-or-face-legal-action>. (última consulta, 5 de diciembre del 2019)

_____. “Facebook, Twitter, Microsoft, YouTube Launch Shared Terrorist Media Database”. *Wired UK*. Disponible en la siguiente página web: <http://www.wired.co.uk/article/facebook-twitter-microsoft-youtube-launch-shared-terrorism-database>. (última consulta, 6 de diciembre de 2019)

_____. “UK Government Wants ‘Unsavoury’ Web Content Censored”, *Wired UK*. Disponible en la siguiente página

web: <http://www.wired.co.uk/article/government-web-censorship>. (última consulta, 15 de marzo de 2019)

CRAIG, SCOTT, y LLANSÓ, EMMA. “Pressuring Platforms to Censor Content is Wrong Approach to Combatting Terrorism”. *Center for Democracy & Technology*. Disponible en la siguiente página web: <https://cdt.org/blog/pressuring-platforms-to-censor-content-is-wrong-approach-to-combatting-terrorism/>. (última consulta, 5 de noviembre de 2019)

COMISIÓN EUROPEA. “La Comisión Europea y las compañías de TI anuncian un Código de conducta sobre el discurso ilegal de odio en línea”, Comunicado de prensa IP/16/1937. Disponible en la siguiente página web: http://europa.eu/rapid/press-release_IP-16-1937_en.htm. (última consulta, 31 de mayo de 2019)

_____. “Foro de Internet de la UE: reunir a gobiernos, Europol y compañías de tecnología para contrarrestar el contenido terrorista y el discurso de odio en línea”, Comunicado de prensa IP/15/6243. http://europa.eu/rapid/press-release_IP-15-6243_en.htm. (última consulta, 3 de diciembre de 2019)

_____. “Foro de Internet de la UE: un gran paso adelante para frenar el contenido terrorista en Internet”, Comunicado de prensa IP/16/4328. Disponible en la siguiente página web: http://europa.eu/rapid/press-release_IP-16-4328_en.htm. (última consulta, 8 de diciembre de 2019)

CONSEJO DE EUROPA. Disponible en la siguiente página web: <https://www.coe.int/en/web/tbilisi/-/council-of-europe-secretary->

general-concerned-about-internet-censorship-rules-for-blocking-and-removal-of-illegal-content-must-be-transparent-and-prop. (última consulta, 1 de junio de 2019)

_____. Disponible en la siguiente página web: <https://www.coe.int/en/web/tbilisi/-/council-of-europe-secretary-general-concerned-about-internet-censorship-rules-for-blocking-and-removal-of-illegal-content-must-be-transparent-and-prop>. (última consulta, 1 de junio de 2019)

CORTE DE APELACIONES DE LOS ESTADOS UNIDOS. Séptimo Circuito. Caso Fairley vs. Andrews. N.º 578 F.3d., 20 de agosto de 2009.

COUNTER TERRORISM INTERNET REFERRAL UNIT. “Estadísticas de CTIRU, 250.000 piezas de material extremista/terrorista en línea que se eliminará”. Policía de Gran Bretaña. Disponible en la siguiente página web: <http://news.met.police.uk/news/250000thpiece-of-online-extremist-slash-terrorist-material-to-be-removed-208698>. (última consulta, 2 de diciembre de 2019)

EDDY, MELISSA, y SCOTT, MARK. “Delete Hate Speech or Pay Up, Germany Tells Social Media Companies”, *New York Times*. Disponible en la siguiente página web: <https://www.nytimes.com/2017/06/30/business/germany-facebook-google-twitter.html>. (última consulta, 30 de junio de 2019)

EUROPEAN DIGITAL RIGHTS. “Entrada en Derechos Humanos y prevención contra el extremismo violento”. Disponible en la siguiente página web: <http://edri.org/files/2016-UN-consultation.pdf>. (última consulta, 18 de marzo de 2019)

EUROPOL. “EU Internet Referral Unit: Year one Report Highlights”, N.º 1, 2016. Disponible en la siguiente página web: https://www.europol.europa.eu/sites/default/files/documents/eu_iru_1_year_report_highlights.pdf.

FACEBOOK. “Partnering to Help Curb Spread of Online Terrorist Content”, *Facebook Newsroom*. Disponible en la siguiente página web: <http://newsroom.fb.com/news/2016/12/partnering-to-help-curb-spread-of-online-terrorist-content/>. (última consulta, 5 de diciembre de 2019)

_____. “Informe de solicitudes del gobierno: preguntas frecuentes”, *Facebook Newsroom*. 2016. Disponible en la siguiente página web: <https://govtrequests.facebook.com/faq/>.

FREEDOM ONLINE COALITION. “Presentación al Relator especial de la ONU, David Kaye. Estudio sobre la libertad de expresión y el sector privado en la edad digital”, 2016. Disponible en la siguiente página web: https://www.freedomonlinecoalition.com/wp-content/uploads/2016/02/FOC-WG3-Submission_ICT-Sector-Report.pdf.

FRISCHMANN, BRETT, y SELINGER, BRETT. “Ser humano en el siglo XXI”, 2018. (manuscrito).

GUYNN, J. “Twitter to Police Abuse in Major Shift”, 2017. Disponible en la siguiente página web: <https://www.usatoday.com/story/tech/news/2017/03/01/twitter-to-police-abuse-in-major-shift/98559482/>. (última consulta, 2 de diciembre de 2019)

HALLIDAY, J., y WANG, T. “Somos el ala de la libertad de expresión de la fiesta de la libertad de expresión”, *The Guardian*. Disponible en la siguiente página web: <https://www.theguardian.com/media/2012/mar/22/twittertony-wang-free-speech>. (última consulta, 2 de diciembre de 2019)

HERN, ALEX. “Facebook lanza herramientas para enfrentar la venganza porno”, *The Guardian*. (última consulta, 5 de abril de 2019) (digital)

JEONG, SARAH. “Terror Scanning Database For Social Media Raises More Questions than Answers”. *Motherboard*. https://motherboard.vice.com/en_us/article/social-media-terror-scanning-database. (última consulta, 9 de diciembre de 2019)

KAUFFMAN, GRETTEL. “La UE insta a los gigantes de las redes sociales a actuar en el discurso del odio”. *Christian Science Monitor*. Disponible en la siguiente página web: <http://www.csmonitor.com/Technology/2016/1205/EU-urgessocial-media-giants-to-act-on-hate-speech>. (última consulta, 5 de diciembre de 2019)

KNIBBS, KATE. “Extreme Moderation”, *Ringer*. Disponible en la siguiente página web: <https://theringer.com/curbing-terrorist-social-media-activity-facebook-twitter-google-601ff9684068>. (última consulta, 17 de enero de 2019)

LEE, T. “YouTube Rebuffs Senator’s Demands to Remove Islamist Videos”, *Ars Technica*. 2008. (digital)

LEVIN, SAM. “Facebook bloquea temporalmente al activista de Black Lives Matter luego de publicar correos racistas”, *The Guardian*. Disponible en la siguiente página web: <http://www.theguardian.com/technology/2016/sep/12/facebook-blocks-shaun-king-black-lives-matter>. (última consulta, 12 de septiembre de 2019)

LIEBERMAN, J. “Lieberman a YouTuber: eliminar videos de Al-Qaeda”. *CNN*. 2008.

LOVITT, LIANE. “¿Por qué los informes de transparencia importan ahora más que nunca?”. *Medium*. Disponible en la siguiente página web: <http://medium.com/inflection-points/why-transparency-reports-matter-now-more-than-ever-9fb6ebe733fa#yhossoi1q>. (última consulta, 13 de mayo de 2019)

MCGOOGAN, CARA. “EU acusa a Facebook y Twitter de fallar remover el discurso de odio”, *The Daily Telegraph*. Disponible en la siguiente página web: <http://www.telegraph.co.uk/technology/2016/12/05/eu-accusesfacebook-twitter-failing-remove-hate-speech/>. (última consulta, 5 de diciembre de 2019)

MOODY, GLYN. “Internet Giants Will Join Forces to Stop Online Sharing of Terrorist Material”, *ARS Technica*. Disponible e la siguiente página web: <http://arstechnica.com/tech-policy/2016/12/twitter-facebook-microsoft-youtube-terrorist-material-removal/>. (última consulta, 6 de diciembre de 2019)

MILLS, CURT. “Facebook está buscando un analista de antiterrorismo”, *U.S. News & World Report*. <https://www.usnews.com/news/national-news/articles/2016-11-14/facebook-is-looking-for-a-counterterrorism-analyst>. (última consulta, 14 de noviembre de 2019)

MURDOCK, JASON. “ISIS: Policía Cibernética del Reino Unido incrementa la campaña para frenar la propagación de la propaganda inspirada en Daesh”. Disponible en la siguiente página web: <http://www.ibtimes.co.uk/isis-ukcyber-cops-ramp-campaign-curb-spread-daesh-inspired-propaganda-1555463>. (última consulta, 18 de abril de 2019)

NAKASHIMA, ELLEN. “There’s a New Tool to Take Down Terrorism Images on Line. But Social-Media Companies Are Wary of It”, *The Washington Post*. https://www.washingtonpost.com/world/national-security/new-tool-to-take-down-terrorism-images-online-spurs-debateon-what-constitutes-extremist-content/2016/06/20/0ca4f73a-3492-11e6-8758-d58e76e11b12_story.html?utm_term=.28394b560bb2. (última consulta, 21 de junio de 2019)

NEWTON, CASEY. “Facebook, Microsoft, Twitter y YouTube crean una base de datos de contenido terrorista”. Disponible en la siguiente página web: <https://www.theverge.com/2016/12/5/13849570/facebook-microsoft-twitter-google-terrorist-content-database>. (última consulta, 5 de diciembre del 2019)

OFICINA DEL GABINETE Y OFICINA DEL HOGAR. “Plan de Acción Francés-Británico. Seguridad de Internet”, 2017. Disponible en la siguiente página web: <https://www.gov.uk/government/publications/french-british-action-plan-internet-security>.

ORGANIZACIÓN DE LAS NACIONES UNIDAS. Comentarios escritos del Artículo 19 sobre el Informe del Relator Especial sobre la Promoción y Protección del Derecho a la Libertad de Opinión y Expresión. Doc. A/HRC/32/28. <http://www.ohchr.org/Documents/Issues/Expression/PrivateSector/Article19.pdf>. (última consulta, 11 de mayo de 2019)

PAULSON, AMANDA, y BOTKIN-KOWACKI, EVA. “En la lucha contra el terrorismo, compañías tecnológicas atrapadas entre los ideales de Estados Unidos y Europa”. *Christian Science Monitor*. Disponible en la siguiente página web: <https://www.csmonitor.com/Technology/2017/0623/In-terror-fight-tech-companies-caught-between-USand-European-ideals>. (última consulta, 23 de junio de 2019)

PÉREZ, SARAH. “Facebook, Microsoft, Twitter y YouTube colaboran para eliminar contenido terrorista de sus servicios”, *Techcrunch*. Disponible en la siguiente página web: <http://techcrunch.com/2016/12/05/facebook-microsoft-twitter-and-youtube-collaborate-to-remove-terrorist-content-from-their-services/>. (última consulta, 5 de diciembre de 2016)

PEYSER, EVE. “Twitter and Facebook Randomly Crack Down on Terrorist Videos After EU Warning”, *Gizmodo*. Disponible en la siguiente página web: <http://gizmodo.com/twitter-and-facebook-decide-to-crackdown-on-spread-of-1789710389>. (última consulta, 5 de diciembre de 2019)

PHELPS-ROPER, MEGAN. “Crecí en la Iglesia Bautista de Westboro. He aquí por qué me fui”, *TED Talk*. Disponible en la siguiente página web: https://www.ted.com/talks/megan_phelps_roper_i_grew_up_in_the_westboro_baptist_church_here_s_why_i_left?utm_campaign=social&utm_medium=referral&utm_source=facebook.com&utm_content=talk&utm_term=global-social%20issues#t-627390. (última consulta, febrero de 2019)

_____. “Si está motivado a odiar, ¿aún puede revertirlo?”, *TED Radio Hour*. Disponible en la siguiente página web: <https://www.npr.org/2017/10/17/560181511/megan-phelps-roper-if-youre-raised-to-hate-can-you-reverse-it>. (última consulta, 27 de octubre de 2019)

POST, ROBERT. “Hate Speech”, Hare, Ivan, y Weinstein, James (eds.). *Extreme Speech and Democracy*. 2009.

RADSCH, COURTNEY C. “Privatizing Censorship in Fight against Extremism Is Risk to Press Freedom”, *Committee to Protect Journalists*. Disponible en la siguiente página web: <https://cpj.org/blog/2015/10/privatizing-censorship-in-fight-against-extremism-.php>. (última consulta, 16 de octubre de 2019)

SCOTT, MARK. “Europa presiona a las empresas estadounidenses de tecnología para abordar el discurso del odio”. *New York Times*. Disponible en la siguiente página web: <https://www.nytimes.com/2016/12/06/technology/europe-hate-speechfacebook-google-twitter.html>. (última consulta, 6 de diciembre del 2019)

SHACKFORD, SCOTT. “How Long Before This Tool to Censor Images from Terrorists Gets Misused?”, *Reason*. Disponible en la siguiente página web: <http://reason.com/blog/2016/12/06/how-long-before-thistool-to-censor-image>.(última consulta, 6 de diciembre de 2016)

TOOR, AMAR. “Legisladores del Reino Unido dicen que Facebook, Google y Twitter están ‘constantemente fallando’ en la lucha contra ISIS en línea”. Disponible en la siguiente página web: <http://www.theverge.com/2016/8/26/12656328/facebook-google-twitter-isis-propaganda-uk-report>. (última consulta, 26 de agosto de 2019)

_____. “Facebook, Twitter, Google, and Microsoft Agree to EU Hate Speech Rules”, *Verge*. <http://www.theverge.com/2016/5/31/11817540/facebook-twitter-googlemicrosoft-hate-speech-europe>. (última consulta, 31 de mayo de 2019)

TWITTER. “Informes TDS al gobierno”. Disponible en la siguiente página web: <https://transparency.twitter.com/en/gov-tos-reports.html>. (última consulta, 9 de julio de 2019)

_____. “Twitter Trust and Safety Council”. Disponible en la siguiente página web: <https://about.twitter.com/safety/council>. (última consulta, 29 de septiembre de 2019)

UNIÓN EUROPEA. “Código de conducta para contrarrestar el discurso de odio ilegal en línea”, 2016. Disponible en la siguiente página web: https://ec.europa.eu/newsroom/document.cfm?doc_id=42855. (última consulta, 5 de octubre de 2019)

_____. Código de conducta para contrarrestar el discurso de odio ilegal en línea. https://ec.europa.eu/newsroom/document.cfm?doc_id=42855. (última consulta, 5 de diciembre de 2019)

_____. “Código de Conducta para combatir el discurso ilegal del odio en línea: primeros resultados en la implementación”, Comunicado Europeo. Disponible en la siguiente página web: http://ec.europa.eu/information_society/newsroom/image/document/2016-50/factsheet-código-conducta-8_40573.pdf. (5 de diciembre de 2019)

_____. “Informe de la Comisión de Subsidiariedad y Proporcionalidad”. Disponible en la siguiente página web: http://ec.europa.eu/dgs/secretariat_general/relations/relations_other/npo/subsidiarity_reports_en.htm. (15 de marzo de 2019)

WADDELL, KAVEH. “A Tool to Delete Beheading Videos Before They Even Appear Online”, *The Atlantic*. Disponible en la siguiente página web: <https://www.theatlantic.com/technology/archive/2016/06/a-tool-to-delete-beheading->

videos-before-they-even-appear-online/488105/. (22 de junio de 2019)

WALKER, KENT. “Cuatro pasos que estamos tomando hoy para combatir el terrorismo en línea”. Google. , <https://blog.google/topics/google-europe/four-steps-were-taking-today-fightonline-terror/>. (última consulta, 18 de junio de 2019)

WAMSLEY, LAUREL. “Austrian Court Rules Facebook Must Delete Hate Speech”, *NPR*. Disponible en la siguiente página web: <http://www.npr.org/sections/thetwo-way/2017/05/08/527398995/austrian-courtrules-facebook-must-delete-hate-speech>. (8 de mayo de 2019)

YORK, JILLIAN C. “European Commission’s Hate Speech Deal With Companies That Will Chill Speech”, *Electronic Frontier Found.* Disponible en la siguiente página web: <https://www.eff.org/deeplinks/2016/06/european-commissions-hate-speech-deal-companies-will-chill-speech>. (última consulta, 3 de junio de 2019)

YOUTUBE. “Políticas sobre contenido violento o explícito de Youtube”. Disponible en la siguiente página web: <https://support.google.com/youtube/answer/2802008?hl=es-419>. (última consulta, 6 de junio de 2019)