Universidad de
SanAndrés

Universidad de San Andrés

Departamento de Derecho
Abogacía

Puesta en cuestión de la libertad de expresión en las redes sociales frente a la proliferación del *hate speech*

Agustina Busso

Legajo: 28023

Mentor: Julio César Rivera

Buenos Aires, julio 2021

Índice

Puesta en cuestión de la libertad de expresión en las redes sociales frente a la proliferación del *hate speech*

Las redes sociales constituyen un espacio ideal para la expresión libre y sin censura estatal. Sin embargo, dado su amplio potencial de difusión - sumado con las notas características del Internet-, dichas plataformas se han convertido en instrumentos ideales para la proliferación del discurso de odio. Dadas las graves consecuencias que dicho fenómeno puede ocasionar en el mundo fuera de línea, el presente trabajo reflexiona acerca de la posible necesidad de una alternativa de regulación para combatir con la difusión de este tipo de discurso a través de las redes sociales. Para ello, luego de un análisis de los posibles esquemas de regulación, se concluye que hoy en día resulta inevitable la necesidad de considerar una alternativa de regulación estatal que aborde la problemática de la proliferación del hate speech en las redes sociales, pero siempre desde una perspectiva que tenga como eje central la protección del derecho fundamental a la libertad de expresión.

1. Introducción

En los últimos años, las redes sociales han adquirido un rol elemental convirtiéndose en importantes herramientas para la comunicación libre y sin censura estatal. Ajenas al control del gobierno, se constituyen en un espacio ideal para la deliberación y el intercambio de ideas. A este fin influyen las notas características de Internet, a través del cual, el alcance de cualquier tipo de contenido tiene un potencial de multiplicación enorme y de ser difundidos casi instantáneamente. Sin embargo, esto que propician las redes sociales encuentra su contracara en su capacidad de convertirse en el espacio ideal para la propagación del discurso de odio[1]. El poder de estas plataformas de generar un daño exponencial cuando son utilizadas como vehículo para la proliferación del discurso del odio, habilitando al mismo adquirir un nivel de difusión a una escala sin antecedentes, ha generado una gran preocupación a nivel internacional dadas las posibles consecuencias derivadas de esto.

[1] Conocido por su término anglosajón como *hate speech*, términos que utilizaremos como sinónimos en el presente trabajo.

Es por esto que las redes sociales han sido el blanco de vastas críticas por no poner ningún tipo de freno a la expansión de mensajes incitadores al odio y la violencia. Por su parte, las compañías privadas se amparan tras el derecho a la libertad de expresión y a presentarse como un ámbito neutral frente al contenido publicado por sus usuarios. Frente a este panorama, se ha empezado a optar tanto a nivel internacional como a nivel nacional, diversos enfoques jurídicos para restringir este tipo de discurso en las redes sociales. Esto conlleva un importante debate en torno a cómo juegan las limitaciones al discurso del odio frente a la libertad de expresión, un derecho fundamental para las sociedades democráticas y plurales. Habiendo dicho esto, como sabemos, si bien la libertad de expresión es un derecho fundamental, este no es absoluto[2] sino que se encuentra sujeto al ejercicio de otros derechos fundamentales, y frente al estado de la cuestión actual, cabe reflexionar si resulta necesario algún tipo de regulación del discurso de odio *online*.

1.1 La libertad de Expresión: su justificación y el alcance de su protección

La libertad de expresión tiene como finalidad reconocer la inmunidad jurídica de la expresión de las personas, en específico de ciertas expresiones que pueden resultar ofensivas, molestas o dañinas y no solo los discursos que puedan ser reconocidos como políticamente correctos o socialmente aceptables[3]. En esta línea, por el mero hecho de que un mensaje ataque

[2] Tórtora Aravena explica que los derechos fundamentales, tal como lo es la libertad de expresión, son "un conjunto de atributos, cuyo respeto y protección son una de las claves más importantes para evaluar la verdadera legitimidad de un modelo político y social [...] No obstante lo anterior,[…] no son absolutos ni ilimitados, sino que en verdad se encuentran sometidos a una serie de restricciones o limitaciones que provocan que su titular no pueda ejercer válidamente una determinada prerrogativa en ciertas circunstancias" (Tórtora Aravena 2010, 168).

[3] En el caso Fressoz and Roire (1999) v. France, el Tribunal Europeo de Derechos Humanos sostuvo que "la libertad de expresión constituye uno de los pilares fundamentales de una sociedad democrática. Sin perjuicio de lo dispuesto en el párrafo 2 del artículo 10, es aplicable no sólo a las "informaciones" o "ideas" que se reciban favorablemente o se consideren inofensivas o indiferentes, sino también a las que ofenden, conmocionan o perturban. Tales son las exigencias del pluralismo, la tolerancia y la amplitud de miras sin las cuales no existe una sociedad democrática [traducción propia]"

o desprecie ciertas ideas o símbolos, más allá de lo repudiable que nos pueda parecer, se está ejerciendo la libertad de expresión dentro de su ámbito de protección (Teruel Lozano, 2017) .

Tal como explica Rosenfeld (2003), son cuatro las principales justificaciones filosóficas de la libertad de expresión y cada una le atribuye un alcance diferente a su legitimidad. Por su parte, cada una de ellas puede conducir a establecer diferentes límites entre el discurso protegido por este derecho fundamental y el discurso que puede ser restringido sin violar la libertad de expresión. En primer lugar, como explica Rosenfeld, la justificación basada en la democracia[4] plantea que la libertad de expresión deviene fundamental en el proceso de autogobierno democrático, tarea que no se podría llevar a cabo sin poder transmitir y recibir libremente ideas. Este es el cauce para la formación de la opinión pública libre y en consecuencia del pluralismo político (Carrillo Donaire 2015). Por ende, basándonos en esta justificación si bien es necesario proteger el discurso político, "el discurso antidemocrático en general, y el discurso de odio y extremista político en particular, con toda probabilidad no sirven para ningún propósito útil y, por lo tanto, no justificarían protección [traducción propia]" (Rosenfeld 2003, 1533)[5]. Por su parte, en segundo lugar, nos encontramos con la justificación basada en la teoría del contrato social, bajo la cual las instituciones políticas deben ser justificables. Según esta, si bien existe la necesidad de un intercambio y discusión de ideas libre, "no puede excluir ex ante ningún punto de vista que, aunque sea incompatible con la democracia, pueda ser relevante para la decisión de un contratista social de adoptar las

[4] La Corte Europea de Derechos Humanos ha sostenido en reiteradas ocasiones que "la libertad de expresión constituye uno de los pilares fundamentales de una sociedad democrática y una de las condiciones básicas para su progreso y la realización personal de cada individuo [traducción propia]" (Hertel v. Switzerland 1998; Surek v. Turkey, 1999; Ligens v. Austria, 1986).

[5] Sobre este punto, Alkiviadou sostiene que "el discurso de odio es una amenaza para el buen funcionamiento de una sociedad democrática y una fuerza condenatoria para valores centrales como el respeto y la solidaridad (...) En un mundo de creciente populismo y extremismo de derecha, el discurso de odio, como consecuencia de tales fenómenos, debe abordarse seriamente [traducción propia]" (Alkiviadou 2018, 203).

instituciones fundamentales de la política o de aceptar cualquier forma particular de organización política [traducción propia]" (Rosenfeld 2003, 1533). En tercer lugar, nos encontramos con la justificación que se basa en la búsqueda de la verdad adoptada por el juez norteamericano Holmes, y así pasándose a conocer como la justificación basada en el libre mercado de ideas[6]. La última se asienta sobre la idea de que hay más probabilidades de que prevalezca la verdad a través de una discusión libre y abierta (por más que en ella se promuevan falsedades) que a través de cualquier otro medio en el cual se quieran eliminar las falsedades por completo. Por último, el autor señala que según la cuarta justificación la autonomía individual y el respeto requieren la protección de la autoexpresión, por lo cual "todo tipo de expresiones posiblemente vinculadas a la necesidad sentida por un individuo de autoexpresión deben recibir protección constitucional [traducción propia]" (Rosenfeld 2003, 1535). Como podemos intuir, esta justificación es la que ofrece el alcance más amplio de justificación de todo tipo de discurso. Sin embargo, si nos paramos desde un punto de vista menos individualista que no se centra exclusivamente en el hablante ("speaker") sino también en el oyente, entonces el discurso del odio podría ser limitado cuando el mismo pueda socavar la autonomía y el respeto por sí mismos de aquellos a quienes va dirigido.

Más allá del rol fundamental de la libertad de expresión que se desprende de las justificaciones mencionadas precedentemente, esto no implica asumir que sea un derecho absoluto[7]. Distintas cartas de derecho fundamentales que proclaman la libertad de expresión

[6]La metáfora del libre mercado de ideas proviene de la jurisprudencia norteamericana, más precisamente del caso de la Corte Suprema de los Estados Unidos "Abrams v. Estados Unidos" (1919), en el cual, en su voto disidente, el juez Holmes sostuvo que "cuando los hombres se han dado cuenta de que el tiempo ha trastornado muchas creencias combativas, pueden llegar a creer, incluso más de lo que creen en los propios fundamentos de su conducta, que el bien final deseado se alcanza mejor mediante el libre comercio de ideas, que la mejor comprobación de la verdad es el poder del pensamiento para hacerse aceptar en la competencia del mercado [traducción propia]" (Holmes 1919).

[7] *Ibid*., 2.

contemplan la necesidad de limitar y establecer ciertos parámetros para restringir este derecho en salvaguarda de otros intereses públicos o privados. En este sentido, por ejemplo, el Pacto Internacional de Derechos Civiles y Políticos ("PIDCP") establece en su artículo 19 que "toda persona tiene derecho a la libertad de expresión [...] [pero] puede estar sujeto a ciertas restricciones que deberán, sin embargo, estar expresamente fijadas por la ley y ser necesarias para: a) Asegurar el respeto a los derechos o a la reputación de los demás; b) La protección de la seguridad nacional, el orden público o la salud o la moral públicas"(1976)[8]. Por su parte, la Convención Europea de Derechos Humanos ("CEDH") en su art.10[9] establece que la libertad de expresión puede estar sujeta a restricciones "que prescriba la ley y sean necesarias en una sociedad democrática, en interés de la seguridad nacional, territorial integridad o seguridad pública, para la prevención de desórdenes o delitos, para la protección de la salud o la moral (...)". Por último, la Convención Americana sobre Derechos Humanos en su art. 13 sostiene que "toda persona tiene derecho a la libertad de pensamiento y de expresión [...] [pero] estará prohibida por ley [...] toda apología del odio nacional, racial o religioso que constituyan incitaciones a la violencia o cualquier otra acción ilegal similar contra cualquier persona o grupo de personas, por ningún motivo, inclusive los de raza, color, religión, idioma u origen nacional" (1969). Así, los diferentes regímenes reconocen bienes jurídicos tales como la

[8] Asimismo, el artículo 20 del PIDCP también se relaciona a los posibles límites del discurso de odio al establecer que "1. Toda propaganda en favor de la guerra estará prohibida por la ley 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" (PIDCP 1976).

[9] En lo que respecta a los límites del derecho a la libertad de expresión, también resulta importante tener en cuenta el art.17 de la CEDH el cual establece la cláusula de la prohibición de abuso del derecho. Tal como explica Díez Bueso, cuando el TEDH se enfrenta a un conflicto donde se implica el discurso y otro derecho, este puede aplicar el art.17 y considerar que la expresión queda excluida de la protección del art. 10.1. O bien, puede considerar que la expresión se encuentra protegida por el art 10.1 de la CEDH y en este sentido analizar si una restricción al mismo es legitima o no de acuerdo a los parámetros del art. 10.2. Para un estudio más en detalle de cómo ha aplicado estos dos artículos el TEDH, ver: Diez Bueso (2020), "Discurso de odio en las redes sociales: la libertad de expresión en la encrucijada".

seguridad nacional, la prevención del delito, la protección de la moral, de la salud, de la propia imagen y de la reputación -entre otros- que juegan un rol fundamental a la hora de delimitar el derecho a la libertad de expresión.

A su vez, resulta interesante que diferentes legislaciones regulan la libertad de expresión y sus limitaciones configurándose sobre diferentes teorías que la justifican. Por ejemplo, mientras que los Estados Unidos se basan sobre todo en el individualismo y el libertarismo, otras naciones continentales, como por ejemplo Alemania, establecen como valor primordial la inviolabilidad de la dignidad humana y el honor personal (Rosenfeld 2003). Esto resulta importante ya que, como veremos más adelante, las diferentes justificaciones de la libertad de expresión pueden llevar a diferentes tratamientos del discurso del odio y limitaciones a dicho derecho.

1.2 La importancia de las redes sociales para la libertad de expresión y su contracara del *hate speech*

Las redes sociales cuentan con un rol fundamental como herramientas que permiten la comunicación y el acceso a información libre y sin control estatal. La libertad en el contenido en ellas no solo provee un espacio ideal para la libre expresión de los individuos sino que las redes se convirtieron en importantes instrumentos para la difusión de contenidos de empresas, partidos políticos, movimientos minoritarios, entre otros actores sociales[10]. Como bien describe Rahul Uttamchandani, "las redes sociales se han convertido en los principales espacios en los que los usuarios pueden expresarse de forma libre sin mayores restricciones editoriales,

[10] Sobre este punto, Wenguang afirma que "Internet hoy en día permea profundamente la vida pública y los intermediarios ingenian prácticamente todas nuestras comunicaciones y conductas en línea. Los cambios tecnológicos y el desarrollo de Internet desencadenaron la revolución en la infraestructura de la libre expresión. En cierto sentido, las plataformas privadas que crearon y controlan esa infraestructura se están convirtiendo en los `Nuevos Gobernadores´ del habla" (Wenguang 2018, 352).

pudiendo publicar desde sus opiniones y preferencias hasta compartir contenidos propios o de terceros, así como pudiendo interaccionar entre sí y generar debates abiertos a cualquier usuario" (Uttamchandani 2020). Es así que estas plataformas gozan de una peculiar capacidad para generar un espacio deliberativo, participativo y democratizador enriqueciendo el pluralismo, la capacidad de informarse y de participar activamente en el debate público. Como explican Isasi y Juanatey, plataformas como Twitter, Facebook e Instagram han adquirido un potencial democratizador al proporcionar un amplio acceso a fuentes de información ajenas al control de los gobiernos y las grandes corporaciones y al constituirse como un espacio óptimo para la deliberación e intercambio de ideas (2017). Dichas particularidades tornan especialmente relevantes en países como Venezuela donde frente al estado de censura impuesta por el gobierno las redes sociales devienen en espacios para denunciar y discutir sobre temas que de otra manera estarían restringidos en la prensa[11].

Sin embargo, este poder de difusión e intercambio que propician las redes sociales generando un medio para que ideas minoritarias se expresen libremente y sin controles frente al poder hegemónico, encuentra su contracara en su "potencial destructivo para sembrar el odio" (Carillo Donaire 2015, 205). Así, las redes sociales se han convertido en el espacio de expresión ideal para la propagación del *hate speech*. Tal como lo describen Isasi y Juanatey, "todas las ideologías intolerantes encuentran en las redes sociales un espacio de expresión privilegiado, que ha generado una especie de cultura del odio, que contaminan e intoxican las redes con lenguaje abusivo, denigrante o agresivo, por motivos, en gran medida, de intolerancia

[11] Tal como explica Goyret (2021), la dictadura de Maduro pretende aumentar los controles en Venezuela para limitar aún más la libertad de expresión de sus ciudadanos a través del seguimiento de las redes sociales, las cuales, tal como explica el autor, "en los últimos años se han convertido en el principal canal al que acuden los venezolanos para informarse, expresarse, e incluso denunciar la dramática situación que atraviesa el país".

contra población inmigrada, refugiados, musulmanes, homosexuales, y otras minorías" (Isasi y Juanatey 2017, 6).

1.3 Las particularidades del discurso de odio *online*

Sumado a todas las aristas que implica debatir sobre los límites de la libertad de expresión frente al discurso del odio, en lo que respecta a este trabajo, nos encontramos con una particularidad adicional: se trata de discurso de odio *online*, en específico, llevado a cabo a través de las redes sociales. Si el discurso del odio tiene un gran potencial de generar daño a sus receptores, Internet y las redes sociales le otorgan un efecto multiplicador que le da al mensaje de odio una capacidad de transmisión que agiganta su potencial dañino.

Otro aspecto de las redes sociales que exacerba los efectos del *hate speech* es el anonimato y el uso de pseudónimos, los cuales generan una sensación de impunidad para continuar difundiendo mensajes de odio (Isasi y Juanatey 2017). A su vez, el hecho de que Internet sea transnacional, tiene como resultado un efecto deshinibidor a la hora de decidir publicar contenidos ya que se dificulta la persecución de quienes lo difunden. Sumado a esto, Internet se ha convertido en un elemento representativo de la liberación y, en consecuencia, cualquier intento de regular lo que suceda en este, y en específico en las redes sociales, sufre de altas probabilidades de ser tachado de ilegítimo y antidemocrático (Isasi y Juanatey 2017).

Sin embargo, más allá de generar un ecosistema favorable para la difusión de mensajes de odio, la realidad es que no hay que pasar por alto que dichas plataformas siguen siendo canales de comunicación de ideas e información y, por ende, merecen la misma protección bajo el alcance de la libertad de expresión que cualquier otro canal para la expresión del discurso (Díaz 2017). En este sentido argumenta Boix Palop quien manifiesta que "puede ocurrir, en efecto, que una expresión realizada en redes sociales logre de facto mayor publicidad y

provoque más daño que si se hubiera realizado por otras vías, pero (...) nada en la expresión en Internet o en redes sociales, en sí misma considerada, debiera hacernos considerar a un mensaje intrínsecamente peor que si es comunicado por otros canales" (Boix Palop 2016, 65). Así, el hecho de que las redes faciliten la propagación de ideas en principio debería verse como algo que contribuye a generar un debate más robusto en donde su alcance permite mayores posibilidades para que diferentes ideas sean conocidas por las personas alrededor del mundo, pudiendo opinar sobre ellas y rebatirlas, cooperando así a la construcción de la opinión pública (Boix Palop 2016).

1.4 Las consecuencias del *hate speech* a través de las redes sociales

Este fenómeno en el cual las redes sociales han contribuido a la proliferación en una escala sin antecedentes del discurso del odio ha generado una gran preocupación a nivel internacional dadas las consecuencias derivadas de este. Si bien resulta esencial que todo tipo de discurso pueda ser expresado de manera libre en vistas a enriquecer la pluralidad de las sociedades democráticas, no resultan menores las repercusiones que puede tener el discurso de odio a una escala tan masiva como la que habilitan las redes sociales.

En este apartado, nos enfocaremos en cómo la proliferación del *hate speech* a través de las redes sociales contribuye a la estigmatización, marginalización y deshumanización de ciertos colectivos históricamente víctimas de abuso y amenazas y que, en última instancia, envían mensajes que segregan y polarizan a la sociedad (Isasi y Juanatey 2017). Esto resulta preocupante cuando tomamos conciencia de que dichos abusos no permanecen en el plano *online* sino que se ven plasmados en actos de violencia en el mundo *offline*. Así, Isasi y Juanatey afirman que mediante la difusión del discurso de odio "se puede generar el caldo de cultivo adecuado para justificar actos discriminatorios, abusos y actos violentos de diversa naturaleza" (Isasi y Juanatey 2017, 9). De esta manera, explican que aunque no sea posible

afirmar de modo general una conexión directa entre la difusión del discurso *online* y los crímenes de odio, resulta cada vez más evidente que existe un vínculo indirecto entre ambos fenómenos. En esta línea, sostienen que torna evidente que los episodios de crímenes violentos de odio rara vez ocurren sin una previa deshumanización y estigmatización de las víctimas (Isasi y Juanatey 2017) [12].

A continuación, a través del análisis de dos ejemplos ilustraremos cómo el discurso de odio *online* puede conducir o desencadenar a actos de violencia más allá de las redes sociales.

a) <u>El genocidio de Myanmar</u>

Un ejemplo paradigmático de este punto es el del crucial papel que tuvo Facebook en el genocidio que se llevó a cabo en Myanmar contra parte de su población musulmana. El conflicto estalló el 25 de agosto de 2017 cuando las Fuerzas Armadas de Myanmar, de mayoría budista, junto con civiles budistas, comenzaron una "campaña de limpieza" contra los rohingyas, un grupo étnico musulmán. La operación militar involucró violaciones, asesinatos masivos, la quema de miles de viviendas, arrestos y detenciones arbitrarias y, en última instancia, obligó a miles de rohingyas a huir a Bangladesh[13].

[12] En el caso Sürek v. Turkey (1999), el Tribunal Europeo de Derechos Humanos sostuvo esta relación entre el discurso de odio y las acciones de odio, explicando que el contenido del discurso: "debe considerarse capaz de incitar a una mayor violencia en la región al inculcar un odio irracional y profundamente arraigado (...) lo que está en juego en el presente caso es el discurso de odio y la glorificación de la violencia [traducción propia]". El caso versaba sobre el dueño de una revista publicada en Estambul a quien dicho Estado demandó por la publicación de dos cartas que, según el demandante, provocaban la enemistad y odio entre la gente. Frente al reclamo del demandado ante el TEDH por la violación de su derecho a la libertad de expresión garantizado por el art. 10 de la Convención, el Tribunal sostuvo que dada la delicada situación de seguridad en el sureste de Turquía y las acusaciones de las cartas que condenaban las acciones militares en dicho territorio, las medidas del gobierno turco fueron tomadas dentro del marco de los objetivos legítimos del segundo párrafo del art. 10 de la Convención dado que fueron en protección de la seguridad nacional, la integridad y prevención de la delincuencia.

[13] Información obtenida de https://www.lanacion.com.ar/el-mundo/la-onu-le-exige-a-myanmar-prevenir-el-genocidio-rohingya-nid2326871/

El reporte llevado a cabo por las Naciones Unidas sobre la Misión Internacional Independiente de Investigación sobre Myanmar realizó un análisis del presunto papel del *hate speech* en el estallido de violencia en Myanmar. En este se informó que no caben dudas de que en Myanmar hay un extendido discurso del odio contra los musulmanes en general y los rohingya en particular, el cual tiene como tema principal el de "amenaza musulmana". En cuanto al rol que jugaron las redes sociales en este contexto, la ONU informa que "la Misión ha sido testigo de una gran cantidad de discursos de incitación al odio en todo tipo de plataformas (...) Los mensajes que describen a los rohingya como violentos, deshonestos, anti-Bamar, anti-budistas, inmigrantes ilegales y / o terroristas [...] están particularmente extendidos en las redes sociales [traducción propia]" (Consejo de Derechos Humanos de la ONU 2018, 340). En este aspecto, dado que Facebook sirvió como la principal plataforma por donde se difundió el odio que alimentó la violencia contra la minoría musulmana, es importante tener en cuenta el distintivo dominio que detenta dicha plataforma en Myanmar. Así, explica el informe de la ONU que previo a 2011, a causa de la censura impuesta por el régimen militar, los medios de comunicación e Internet solo se encontraban disponibles para unos pocos y era extremadamente caro su acceso. Sin embargo, a partir de ese año, mediante un proceso de liberalización de la industria de comunicación, aumentó rápidamente el acceso de la población a Internet y, en específico, a las redes sociales. El reporte comenta que Facebook es la red social más utilizada en Myanmar, a tal punto que "ha llevado a una situación en Myanmar donde Facebook es Internet (...) Es la plataforma principal, si no la única, de noticias en línea (...) En un contexto de escasa alfabetización en medios digitales y sociales, el uso que hace el Gobierno de Facebook para anuncios oficiales y para compartir información contribuye aún más a la percepción de los usuarios de Facebook como una fuente confiable de información [traducción propia]" (Consejo de Derechos Humanos de las Naciones Unidas 2018, 341) . En

consecuencia, podemos ver como esta red social ha contribuido fuertemente a la difusión de la retórica de odio y división.

Ahora bien, a la hora de determinar si las campañas de odio llevadas a cabo por medio de las redes sociales han contribuido o hasta provocado los estallidos de violencia contra los rohingya, para Vuarambon (S.f) resulta claro que la violación de los derechos humanos ocurrida en Myanmar hubiese ocurrido con o sin la existencia de Facebook[14]. Sin embargo, sí se cree que dicha plataforma aumentó la crueldad y la aceptación de parte de la población. El informe apunta, en base a la información que revela el estrecho vínculo entre la incitación al odio en línea y los actos de violencia, que la difusión de *hate speech* en Facebook "contribuyó significativamente a aumentar la tensión y a crear un clima en el que las personas y los grupos pueden volverse más receptivos a la incitación y los llamamientos a la violencia [traducción propia]" (Consejo de Derechos Humanos de las Naciones Unidas 2018, 343)[15]. Por su parte, Facebook admitió que la plataforma fue utilizada para incitar la violencia contra los

[14] Sobre este punto, resulta interesante poder pensar en un paralelismo con el Genocidio de Ruanda (1994) y el papel que jugó la radio en aquel entonces, en específico la Radio-Télévision Libre des Milles Collines ("RTLM"). Tal como explica Kimani (2015), si bien es relevante recordar que las transmisiones de dicha radio no fueron las responsables de introducir la ideología del odio y la polarización existente en la sociedad ruandesa, RTLM fue la estación de radio que sirvió como portavoz de los ideales de la elite Hutu y fue el medio que utilizó para difundir el odio, alimentar un clima de intolerancia y, en definitiva, propagar la guerra contra los Tutsis.

Tal fue el rol que jugó esta radio en la masacre, que en 2003 el Tribunal Penal Internacional de la ONU para Ruanda condenó a Nahimana y Barayagwiza, dado su rol en RTLM, por los crímenes de genocidio, incitación al genocidio, conspiración, crímenes de lesa humanidad, exterminio y persecución. En lo que respecta al poder de la radio -el medio de comunicación con mayor alcance público en su momento (y que podríamos analizar frente al poder de las redes sociales hoy en día)-, el tribunal expresó que RTLM "fue el arma preferida de Nahimana, que utilizó para instigar el asesinato civil tutsi [traducción propia]". A continuación, un link para poder escuchar extractos de transmisiones de RTLM: https://www.youtube.com/watch?v=VNbUeLnxQEI

[15] Explica el reporte que "la Misión recibió información que sugería que el vínculo entre la incitación al odio en línea y fuera de línea y los actos de discriminación y violencia en el mundo real es más que circunstancial. Están surgiendo patrones de sermones y retórica de incitación al odio en lugares específicos que posteriormente han experimentado violencia, y también son indicios de picos de incitación al odio en línea en torno a brotes de violencia [traducción propia]" (Consejo de Derechos Humanos de las Naciones Unidas 2018, 331).

musulmanes en Myanmar. Así, en un artículo de la empresa publicado por Alex Warofka, (*Product Policy Manager* de Facebook), la red social declaró que si bien ellos quieren que la plataforma sea un lugar para que las personas se puedan expresar con libertad y seguridad, Facebook no hizo lo suficiente para evitar ser utilizada como un lugar donde se fomente la división y se incite la violencia fuera de línea.[16]

b) La radicalización *online* - El ataque terrorista en Nueva Zelanda

Otro de los fenómenos desarrollados a partir de las redes sociales y su potencial de difusión ha sido la propaganda terrorista en línea. El Consejo de la Unión Europea ha hecho referencia a esto explicando que la amenaza terrorista se ha desarrollado velozmente en los últimos años dado que han utilizado Internet como medio para inspirar y movilizar a individuos a unirse a redes terroristas. En este sentido, la Decisión explica que Internet ha contribuido a la provocación de delitos de terrorismo, a la captación y entrenamiento de terroristas dado que la plataforma "sirve de fuente de información sobre medios y métodos terroristas, funcionando por lo tanto como un «campo de entrenamiento virtual»" (Decisión Marco 2008/919/JAI). El estudio llevado a cabo por Badawy & Ferrara (2018) es una clara demostración del importante papel que detentan las redes sociales para difundir el mensaje terrorista. Los autores explican que los grupos militantes utilizan Internet y redes sociales como Facebook, Instagram, Tumblr y Twitter para difundir su mensaje y reclutar potenciales militantes. Haciendo referencia especialmente al Estado Islámico de Irak y el Levante ("ISIS"), estos afirman que "ningún grupo hasta la fecha ha sido tan inteligente en términos de su campaña de propaganda y reclutamiento de terroristas a través de Internet, y específicamente a través de plataformas de redes sociales, como el Estado Islámico de Irak y el Levante [traducción propia]" (Badawy & Ferrara 2018, 2).

[16] Ver más en: https://about.fb.com/news/2018/11/myanmar-hria/

Si bien los términos “incitación al terrorismo” y “discurso de odio” no son sinónimos, tal como explica Coche, la superposición entre ambos se explica en el hecho de que el terrorismo se basa en ideologías extremistas ya sea respecto a ideologías religiosas, etnonacionalistas, de izquierda y anarquistas, y en este sentido es que la lucha contra el *hate speech online* sirve a los fines de contrarrestar la radicalización (Coche 2018). A su vez, Julian King, comisario de la Unión Europea, hace referencia a la estrecha relación entre la radicalización como consecuencia del discurso de odio y los actos terroristas explicando que: “existe un vínculo directo entre los recientes ataques en Europa y el material en línea utilizado por grupos terroristas como Daesh para radicalizar a los vulnerables y sembrar el miedo y la división en nuestras comunidades [traducción propia]" (King 2017).[17][18]

Un claro ejemplo del discurso de odio llevado a cabo por supremacistas y radicalizados que desencadenó en hechos de violencia en el mundo *offline* fue el del ataque terrorista ocurrido en marzo de 2019 a dos mezquitas de la localidad de Christchurch en Nueva Zelanda. Lo particular del caso es que el agresor grabó el ataque y lo transmitió en vivo a través de Facebook y si bien los videos fueron eliminados, las copias de este se multiplicaron y difundieron por otras plataformas. El atacante venía expresando sus ideas supremacistas a través de diferentes redes sociales. Así, por ejemplo, este había publicado un manifiesto en Twitter donde explica “por qué ha perpetrado el atentado en Nueva Zelanda (...) [y] asegura que quería crear una

[17] El reporte completo de la Comisión Europea, titulado “ Fighting Terrorism Online: Internet Forum pushes for automatic detection of terrorist propaganda” puede ser consultado en el siguiente link: https://ec.europa.eu/commission/presscorner/detail/en/IP_17_5105

[18] En esta misma línea, Citron sugiere que “con menos propaganda terrorista y menos discursos de odio en línea, podría haber menos personas que se unan a los combatientes de ISIS en Siria o coloquen bombas en mercados comerciales o lugares de culto. La evidencia ha sugerido que los seguidores de ISIS que cometieron atrocidades han utilizado las redes sociales para difundir propaganda terrorista y justificar la violencia [traducción propia]” (Citron 2018, 1049).

atmósfera de miedo e incitar a la violencia contra los musulmanes".[19] En definitiva, podemos ver el estrecho vínculo entre odio supremacista y el *hate speech* y como este encuentra un canal de amplia difusión a través de las redes sociales que, en última instancia, puede desencadenar en actos de violencia en el mundo fuera de línea, en este caso en específico en actos de terrorismo.[20]

En conclusión de este apartado, y luego de analizar los dos ejemplos precedentes, se podría decir que la difusión de discursos de odio en Internet, y en específico en las redes sociales, puede llevar a graves actos de violencia en el mundo *offline*. En base a esto, parecería necesario reflexionar sobre algún tipo de regulación para poder controlar el fenómeno del *hate speech* llevado a cabo a través de dichas plataformas. Sin embargo, a la hora de pensar en formas de limitar el discurso del odio tenemos que hacer especial hincapié en lo delicada y riesgosa que es esta actividad, dado que del otro lado nos encontramos con el ejercicio de un derecho tan fundamental como lo es la libertad de expresión. Así, como bien explica en un comunicado de prensa Věra Jourová, "abordar el discurso del odio en línea es un ejercicio delicado que requiere definir claramente dónde se detiene la libertad de expresión y dónde comienza el discurso del odio. La libertad de expresión es un derecho humano, pero este

[19] Para más información sobre los hechos, acceder al siguiente link: https://miquelpellicer.com/2019/03/apuntes-de-comunicacion-sobre-la-matanza-en-apuntes-de-comunicacion-sobre-la-matanza-en-christchurch/

[20] Así es como Vera Jourová, Comisaria de la Unión Europea, explica que "los recientes ataques terroristas nos han recordado la urgente necesidad de abordar el discurso de odio ilegal en línea. Las redes sociales son, lamentablemente, una de las herramientas que utilizan los grupos terroristas para radicalizar a los jóvenes y uso racista para difundir la violencia y el odio [traducción propia]" (Jourová 2016).

Acceder al siguiente link para poder ver el reporte completo de la Comisión Europea, titulado: "European Commission and IT Companies announce Code of Conduct on illegal online hate speech" https://ec.europa.eu/commission/presscorner/detail/en/IP_16_1937

derecho no protege el discurso de odio ilegal que incita a la violencia y al odio" (Jourová 2016).[21]

2. La regulación del *hate speech*

En vistas a la dimensión que ha adquirido el discurso del odio en internet y, en específico, a las consecuencias que trae aparejado en el mundo *offline*, es menester reflexionar sobre la necesidad de algún tipo de regulación para contrarrestar los efectos de este tipo de discurso. De hecho, esta cuestión se encuentra cada vez más en las agendas de los Estados y diferentes organismos internacionales. Sin embargo, encontrar el balance entre la demarcación de los límites del *hate speech* y la protección de la libertad de expresión es una tarea extremadamente complicada, sino riesgosa para el derecho fundamental que está en juego. Tal como explica Mertsching (2018), si bien la inexistencia de regulación de este tipo de discurso puede tener graves consecuencias, las iniciativas para regular el discurso del odio en las redes sociales suelen ser controvertidas dado que " regular cualquier forma de discurso generalmente pone en juego el derecho fundamental a la libertad de expresión (...) Por un lado, deben evitarse los contenidos ilegales; Sin embargo, al fin y al cabo, esto no debería evitar una comunicación legal o incluso socialmente valiosa"(Mertsching 2018, 1).

A la hora de pensar en maneras de regular este tópico, se nos presentan al menos tres perspectivas desde las cuales podría partir la regulación del discurso de odio en redes sociales. La primera forma de hacerlo es a través de la relación individuo- Estado. En este sentido, a través de la regulación estatal se puede imponer sanciones tanto de tipo penal como civiles a

[21] Acceder al siguiente link para poder ver el reporte completo de la Comisión Europea, titulado: "EU Internet Forum: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online" https://ec.europa.eu/commission/presscorner/detail/en/IP_15_6243

los usuarios de las redes sociales que difundan a través de Internet determinadas expresiones que puedan ser definidas como discurso de odio.

En segundo lugar, otra relación que se puede abordar a la hora de reflexionar acerca de la regulación del *hate speech* es la de red social-usuario. Acá hacemos referencia en principio a las propias reglas establecidas por la red social a través de las cuales puede decidir eliminar ciertos mensajes creados por los usuarios o hasta eliminar su cuenta en tanto mediante su contenido estos violen sus términos y políticas de servicio. En este trabajo, se hará referencia a los términos y condiciones establecidos por las redes sociales para regular el *hate speech* como "regulación privada".

En tercer lugar, la relación sobre la que se puede reflexionar y la que este trabajo abordará en profundidad es la de Estado-red social. En este sentido, analizaremos hasta qué punto pueden los Estados, tanto a nivel local como internacional (a través de Convenciones Internacionales u otros instrumentos internacionales), obligar a las redes sociales a ejercer control sobre el contenido generado por sus usuarios bajo apercibimiento de sancionarlas en caso de no cumplir con dichos controles. Este tipo de regulación será denominada para el presente trabajo como "regulación pública/estatal".

Trataremos esta relación en particular por varias razones. Por un lado, porque la libertad de expresión como derecho fundamental es reconocida a los ciudadanos frente a los poderes públicos y no en relación a los entes de carácter privado, por lo cual en principio se podría pensar que las redes sociales pueden poseer sus propias políticas de contenido sin esto afectar la libertad de expresión de los usuarios [22]. Asimismo, porque tal como afirma Wenguang hoy

[22] Para sostener esto, me apoyo en la concepción estadounidense en materia de derechos individuales y su desconfianza frente al Estado. Al respecto, Rivera explica que "el derecho constitucional estadounidense está sustentado en la idea de que el poder estatal es categóricamente más peligroso para la libertad individual que el poder privado. Por lo tanto, los derechos reconocidos en la Constitución protegen al individuo exclusivamente frente al Estado, pero no frente a otros individuos [...] De esta

en día “las plataformas son los campos de batalla centrales sobre la libertad de expresión en la era digital. Teniendo en cuenta la prevalencia del discurso de odio en línea y sus daños a las personas objetivo, el discurso democrático y la seguridad pública, es necesario combatir el discurso de odio en línea. Para ello, los intermediarios juegan un papel crucial como los nuevos gobernantes del discurso en línea [traducción propia]” (Wenguang 2018, 356). En este sentido, hoy en día parece esencial reflexionar sobre una regulación que tenga como sujeto la plataforma que sirve como medio para la difusión de los contenidos que nos preocupan. Además, teniendo en cuenta las características particulares de este medio, tales como el anonimato, la transnacionalidad y la capacidad de difundir contenidos exponencialmente por medio de diferentes plataformas (y en caso de que te bloqueen el contenido en una rápidamente la puedes difundir en otra), la realidad es que ya no parece el medio más eficiente perseguir a los individuos generadores de contenido -pareciendo una tarea casi imposible tener el control sobre los miles de millones de usuarios de redes sociales[23]- sino que deviene mucho más eficaz pensar en una regulación directamente para la plataforma que habilita la creación y difusión del contenido en sí. Así, “se plantea que quizás sea más efectivo si no se puede perseguir a los autores del discurso de odio quitarles su altavoz, es decir, tomar medidas dirigidas a los intermediarios de Internet” (Gascón Marcen 2019, 65). Por último, encarar el trabajo desde un

manera, las conductas de sujetos no estatales que afectan los derechos o intereses de otros sujetos no estatales no constituyen una cuestión constitucional puesto que los derechos constitucionales sólo pueden ser opuestos frente al Estado” (Rivera 2006, 7).

Resulta relevante destacar que hay quienes defienden una postura contraria a lo susodicho. Así, nuestra Corte Suprema de Justicia en el caso KOT, Samuel S.R.L. s/ Acción de amparo, en su voto mayoritario sostuvo “si bien en el precedente citado la restricción ilegítima provenía de la autoridad pública y no de actos de particulares, tal distinción no es esencial a los fines de la protección constitucional [...] Nada hay, ni en la letra ni en el espíritu de la Constitución, que permita afirmar que la protección de los llamados "derechos humanos" - porque son los derechos esenciales del hombre- esté circunscripta a los ataques que provengan sólo de la autoridad”.

[23] Según el reporte llevado a cabo por Hootsuite y We Are Social “Más de 500 millones de nuevos usuarios se unieron a las plataformas de redes sociales durante los últimos 12 meses, lo que eleva el total mundial a 4,33 mil millones en abril de 2021”. Ver más en: https://wearesocial.com/blog/2021/04/60-percent-of-the-worlds-population-is-now-online

enfoque concentrado en la regulación privada lisa y llana hoy en día parecería un tanto *naif* o ingenuo dado que, como veremos en apartados posteriores, actualmente las redes sociales no diseñan sus políticas de manera autónoma sino que es cada vez mayor la presión pública, sobre todo por parte de la Unión Europea, ante la que sucumben las compañías en cuanto al diseño de su regulación interna frente a cómo tratar el contenido generado por sus usuarios.

2.1. La regulación pública del contenido publicado en las redes sociales

El discurso de odio plantea complejidades a las democracias occidentales contemporáneas frente a la cuestión de cómo regularlo y los límites aceptables al derecho fundamental de libertad de expresión. Por eso, en este apartado examinaremos cómo los Estados e instituciones públicas han respondido frente al fenómeno de las redes sociales y el poder que han adquirido para la difusión de todo tipo de información y discursos.

2.1.1. Los dos paradigmas principales de regulación: "De la tolerancia norteamericana a la intransigencia europea"[24]

En cuanto a su tratamiento constitucional en el ámbito internacional, si bien está lejos de ser uniforme, se suelen reconocer dos grandes posturas. Por un lado, la perspectiva estadounidense que se caracteriza por una regulación mínima y tolerante al discurso del odio el cual se encuentra ampliamente amparado por la Primera Enmienda de la constitución norteamericana. Por el otro, el modelo europeo y de los pactos internacionales, tradicionalmente con una regulación mucho más extensa y restrictiva a la libertad de expresión frente a este tipo de contenido, en vistas a la protección de otros derechos y valores que entran en conflicto.

[24] Tal como lo caracteriza Carillo Donaire

Por su lado, la tradición norteamericana ha rechazado las restricciones estatales a las expresiones de odio basándose en la protección a ultranza de la Primera Enmienda, la cual predica que "el Congreso no promulgará ninguna ley con respecto a [...] restringir la libertad de expresión o de prensa" (Primera Enmienda de la Constitución de Estados Unidos 1791). Así, conducidos por la constitución, el ciberliberalismo y el énfasis a promover la industria del Internet, el modelo norteamericano se caracteriza por fomentar amplias exenciones de responsabilidad a los intermediarios en cuanto al contenido ilegal generado por terceros (esto es, los usuarios de las redes sociales) (Wenguang 2018). Esta postura se ha articulado en una concepción que enfatiza el valor del debate público y del libre mercado de ideas, en el que la verdad siempre acaba obteniendo aceptación (Isasi y Juanatey 2017).

Por el contrario, el paradigma europeo ha seguido una línea mucho más intolerante al *hate speech* y restrictivo de la libertad de expresión, brindándole mayor peso a otros derechos y valores tales como el honor, la igualdad, la no discriminación y la libertad religiosa (Carrillo Donaire 2015). Como explica Carrillo Donaire (2015), esta postura intransigente se debe, en gran medida, a la influencia de la historia vivida por Europa en el Siglo XX, en la que expresiones de odio y discursos discriminatorios confluyeron en el holocausto judío bajo el régimen Nazi. Asimismo, el autor agrega que "también contribuye al contraste entre ambos paradigmas la denominada cultura europea del honor, que se asienta en un mayor protagonismo de los valores del honor, la intimidad y la propia imagen en el viejo continente" (Carrillo Donaire 2015, 2019). Es en esta línea que, en los últimos años, Europa ha incrementado su vigilancia hacia los denominados "gigantes tecnológicos de Silicon Valley[25]" a partir de

[25] Silicon Valley es una región ubicada en California que constituye la meca tecnológica y emprendedora más importante del mundo. Las más importantes empresas tecnológicas han establecido sus sedes en Silicon Valley, entre ellas, Google, Facebook y Twitter. Para más información, visitar: https://www.iprofesional.com/tecnologia/308750-que-es-silicon-valley-y-donde-queda-la-meca-de-la-tecnologia

diferentes iniciativas gubernamentales a través de las cuales intentan regular el discurso de odio en Internet.

A continuación, en aras de profundizar en este análisis comparativo de ambos enfoques, analizaremos algunos ejemplos de legislación que ilustran los paradigmas adoptados por Estados Unidos y la Unión Europea.

a) <u>Sección 230 de la *Communications Decency Act* – Estados Unidos</u>

En cuanto al paradigma estadounidense, la sección 230 de la *Communications Decency Act* (en adelante, "CDA") es la expresión misma de la tolerancia estatal frente al discurso del odio[26]. La CDA, norma federal en Estados Unidos, establece que "ningún proveedor o usuario de un servicio informático interactivo será tratado como el editor o hablante de cualquier información proporcionada por otro proveedor de contenido de información" (*Communications Decency Act* 1996). Dicha norma se promulgó en los primeros días de Internet, originalmente con el propósito de promover el mercado de Internet emergente y la innovación en este, y mediante la misma los proveedores de redes sociales gozan de una amplia inmunidad frente a la responsabilidad por los contenidos generados por terceros (Isasi y Juanatey 2017). De esta manera, la misma se ha convertido en una pieza fundamental para los intermediarios en Internet dado que les da dos ventajas. Por un lado, las exime de responsabilidad por el contenido que los usuarios de las redes sociales publiquen. Por el otro, les da la libertad de dar de baja

[26] La jurisprudencia de la Corte Suprema de los Estados Unidos ha demostrado esta misma línea de pensamiento. Así, en el reciente caso "Matal v. Tam", el juez Alito sostuvo que "el Gobierno tiene interés en evitar que los discursos expresen ideas ofensivas. Y, como hemos explicado, esa idea golpea el corazón de la Primera Enmienda. El discurso que degrada por motivos de raza, etnia, género, religión, edad, discapacidad o cualquier otro motivo similar es odioso; pero lo que más nos enorgullece de nuestra jurisprudencia sobre la libertad de expresión es que protegemos la libertad de expresar el pensamiento que odiamos [traducción propia]" (2017).

contenidos e intervenir en lo que se publica en dichas plataformas[27]. Así explica Sevanian que "la mayoría de los tribunales que aplican la Sección 230 han interpretado desde entonces esta ley del "buen samaritano", apropiadamente titulada, como una concesión de inmunidad general a [los intermediarios][28], que ofrece protección independientemente de si un [intermediario] realmente regula o edita su sitio web" (Sevanian 2014, 121).

A modo de ilustración de la importancia de la llamada Sección 230 en la comunicación en Internet tal como la conocemos hoy día, la organización *Electronic Frontier Foundation* explica que "este marco jurídico y político ha permitido que los usuarios de YouTube y Vimeo suban sus propios vídeos, que Amazon y Yelp ofrezcan innumerables opiniones de usuarios, que Craigslist albergue anuncios clasificados y que Facebook y Twitter ofrezcan redes sociales a cientos de millones de usuarios de Internet [traducción propia]" (*Electronic Frontier Foundation* s.f). Así dicha organización explica que dada la inmensa cantidad de contenido generado por los usuarios de estas plataformas, sería inviable que las mismas pudieran impedir la existencia de contenidos objetables en sus sitios. Esto en definitiva generaría que ante la posibilidad de enfrentarse a responsabilidades legales, los intermediarios dejarán de alojar contenido de los usuarios o censuren activamente el contenido.[29]

[27] Continúa la sección 230 estableciendo: "Protección para el bloqueo y filtrado del material ofensivo por parte del "buen samaritano" [...] Ningún proveedor o usuario de un servicio informático interactivo podrá ser considerado responsable a causa de cualquier acción tomada voluntariamente y de buena fe para restringir el acceso o la disponibilidad de material que el proveedor o usuario considere obsceno, lascivo, sucio, excesivamente violento, acosador o de cualquier otro modo objetable, independientemente de que dicho material esté constitucionalmente protegido [traducción propia]" (*Communications Decency Act* 1996).

[28] El autor en su texto se refiere a los intermediarios de Internet como "servicios informáticos interactivos".

[29] Por su parte, en esta misma línea de argumentos, podemos remitir a los dichos de Mark Zuckerberg, fundador y Presidente de Facebook. Zuckerberg expresó, frente al Senado de los Estados Unidos, que "gracias a la Sección 230, la gente tiene libertad de usar internet para expresarse (...) Creemos en darle voz a la gente, incluso cuando eso significa defender los derechos de la gente con la que no estamos de acuerdo" (Zuckerberg 2020). Por su parte Jack Dorsey, el CEO de Twitter, alertó sobre la importancia de esta norma para las plataformas y la libertad de expresión en la audiencia explicando que "socavar

b) *Soft Law* en la Unión Europea - El Código de conducta de la Unión Europea para combatir el discurso de odio ilegal en línea

Por su parte, una medida fundamental de la Unión Europea en torno al camino hacia la regulación del discurso de odio se tomó en 2016 cuando, como resultado del Foro Europeo de Internet, la Comisión Europea anunció el Código de Conducta de la Unión Europea para Combatir el Discurso de Odio Ilegal en Línea (en adelante, el "Código de Conducta" o el "Código"). Con el objetivo de contrarrestar el evidente crecimiento del discurso de odio mediante estas plataformas, el Código consiste en un acuerdo firmado conjuntamente con empresas tecnológicas tales como Facebook, YouTube, Microsoft y Twitter[30] en la cual estas se comprometen en la lucha contra la propagación del *hate speech* a través de Internet mediante diversos mecanismos de control y políticas para eliminar esta clase de contenidos de sus plataformas. Así, en un comunicado de prensa, la Comisión Europea explica que las empresas tecnológicas apoyan "el esfuerzo por responder al desafío de garantizar que las plataformas en línea no ofrezcan oportunidades para que el discurso de odio ilegal en línea se propague de forma viral" [traducción propia]" (Comisión Europea 2016).

la Sección 230 resultará en una mayor eliminación del discurso en línea e impondrá graves limitaciones a nuestra capacidad colectiva para abordar el contenido dañino y proteger a las personas en línea"(Dorsey 2020).

Estos dichos se llevaron a cabo durante la segunda comparecencia a la que tuvieron que asistir los directivos de las *big tech* para defender la Sección 230, frente a los deseos de reforma por parte de los legisladores de Estados Unidos. Para más información, visitar: https://www.infobae.com/america/agencias/2020/10/28/twitter-y-facebook-defienden-su-inmunidad-en-internet-antes-de-audiencia-en-senado-de-eeuu/

[30] Tal como informa la Comisión Europea, en el curso de 2018 se sumaron las plataformas Instagram, Snapchat y Dailymotion, en 2019 se unió Jeuxvideo.com y, por último, TikTok anunció su participación en 2020. Para más información, visitar: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theeucodeofconduct

En aras de guiar sus esfuerzos para combatir el discurso de odio *online*, las plataformas se implican a una serie de compromisos, entre los cuales se destacan[31]:

- Desarrollar procesos y capacitar al personal para la efectiva revisión de notificaciones de denuncias sobre discurso de odio para así poder, de ser necesario, bloquear o eliminar el acceso a dicho contenido en menos de 24 horas.

- Evaluar las solicitudes de eliminación en función de sus normas de autorregulación y directrices comunitarias y, solo cuando sea necesario, de las leyes nacionales que transpongan la Decisión marco 2008/913 / JAI[32].

- A través de sus normas, educar a sus usuarios y crear conciencia sobre el contenido de odio prohibido bajo sus términos de servicio y políticas de contenido.
- Fomentar la provisión de avisos cuando un contenido promueve la incitación a la violencia y al odio por parte de expertos, especialmente mediante la asociación con Organizaciones de la Sociedad Civil.
- Acceder a diferentes Organizaciones de la Sociedad Civil y a "reporteros de confianza" en todos los Estados miembros para ayudar a proporcionar avisos de alta calidad en sus sitios web.

[31] Para información completa de los compromisos asumidos por las plataformas a través del Código de Conducta, ingresar a: https://ec.europa.eu/info/files/code-conduct-countering-illegal-hate-speech-online_en

[32] Esta es la "Decisión marco relativa a la lucha contra determinadas forma y manifestaciones del racismo y la xenofobia mediante el derecho penal" del Consejo de la Unión Europea, cuyo objetivo es que determinadas manifestaciones de racismo y xenofobia sean penalmente punibles en todos los países miembros de la Unión Europea. Para má infromación, acceder a: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=LEGISSUM%3Al33178

- Trabajar para la identificación y promoción de contra-discursos independientes contra la retórica del odio, de nuevas ideas e iniciativas y apoyar programas educativos que fomenten el pensamiento crítico.

Por último, tal como explica la Comisión Europea en su artículo, la implementación de esta herramienta llevada a cabo a través del acuerdo entre la Unión Europea y las empresas tecnológicas se evalúa a través de seguimientos periódicos que se lleva a cabo en colaboración con una red de organización de los diferentes Estados miembros de la Unión. Así, según los reportes que van desde 2016 hasta el 2020, el Código de Conducta ha surtido grandes efectos, ofreciendo una contundente respuesta frente al *hate speech* en las plataformas involucradas[33].

A pesar de esto, hay ciertos aspectos del Código de Conducta que resultan más alarmantes frente al derecho a la libertad de expresión. Uno de los principales puntos al respecto radica en el hecho de que parecería que el Código rebaja a un segundo plano a la ley, tras el papel principal que juegan las compañías privadas proveedoras de las redes sociales (Gascón Marcen 2019). Esto ya que a partir de la implementación arbitraria de sus términos de servicio en relación al *hate speech* podrían terminar eliminando contenido controvertido pero legal, poniendo así en riesgo la libertad de expresión y su protección legal.

[33] De esta manera, tal como informa la quinta evaluación llevada a cabo en junio del 2020, continúa dando resultados positivos con "el 90% de las notificaciones [revisadas] en 24 horas y [eliminando] el 71% del contenido" considerado ilegal como discurso de odio. También llega a la conclusión de que la mayoría de las compañías debe mejorar el *feedback* de las notificaciones recibidas de parte de sus usuarios en general. Así la evaluación menciona que sólo Facebook informa a los usuarios de forma sistemática (el 93,7% de las notificaciones recibieron comentarios). Instagram dio retroalimentación al 62,4% de las notificaciones, Twitter al 43,8% y YouTube solo al 8,8%. Jeuxvideo.com envió comentarios al 22,5% de las notificaciones. Si bien Facebook es la única empresa que informa de manera constante tanto a los notificadores de confianza como a los usuarios en general, Twitter, YouTube e Instagram brindan comentarios con más frecuencia cuando las notificaciones provienen de notificadores de confianza". Para más información, ingresar a: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#monitoringrounds

En conclusión, el Código de Conducta se puede ver como uno de los hitos fundamentales de la Unión Europea y de los Estados que forman parte de esta en el intento por regular y combatir el progresivo discurso de odio que se desarrolla en Internet y en específico en las redes sociales. Dado que los proveedores de redes sociales más importantes y con más influencia han consentido al Código de Conducta, este se convierte en un instrumento legal de gran relevancia (Mertsching 2018). Sin embargo, este no deja de ser un instrumento no vinculante para las redes sociales.

c) Hard Law en la Unión Europea - Ley alemana

Actualmente, a nivel de la Unión Europea no hay ninguna "ley dura" que aborde todas las cuestiones legales relacionadas al discurso de odio *online* de manera satisfactoria[34]. Hoy en día, se podría tomar a la *E-Commerce Directive*[35] como la regulación más importante en este sentido, pero esta no fue diseñada para lidiar con el problema del *hate speech*. Por su parte, como ya mencionamos previamente, si bien el Código de Conducta fue creado específicamente para controlar el discurso del odio en las plataformas sociales, el mismo no reviste carácter vinculante lo cual en definitiva afecta a su aplicabilidad (Mertsching 2018).

[34] De todas maneras, sí hay jurisprudencia del Tribunal Europeo de Derechos Humanos en donde se responsabiliza al intermediario por contenido odioso generado por usuarios. En esta línea, en el caso "Delfi AS v. Estonia" (2015) el tribunal adoptó una postura según la cual los sitios de Internet deben ser considerados responsables por ciertos tipos de comentarios publicados por sus usuarios. Así, el tribunal sostuvo que en casos "donde los comentarios de terceros usuarios se manifiestan en forma de discurso de odio […] los derechos e intereses de los demás y de la sociedad en su conjunto pueden facultar a los Estados contratantes a imponer responsabilidad en los portales de noticias de Internet, sin contravenir el artículo 10 del Convenio, si no toman medidas para eliminar sin demora los comentarios claramente ilícitos, incluso sin previo aviso de la presunta víctima o de terceros [traducción propia]".

[35] Mertsching en este respecto explica que "por ahora, la Unión Europea no proporciona una ley "dura" que aborde explícitamente el discurso de odio. Sin embargo, existen regulaciones para abordar el contenido ilegal en línea, de las cuales algunas son relevantes para el discurso de odio ilegal. Las normas de derecho "duro" más relevantes son la Directiva sobre comercio electrónico, la Directiva sobre terrorismo y la Decisión marco del Consejo 2008/913 / JAI sobre la lucha contra determinadas formas y expresiones de racismo y xenofobia mediante el derecho penal [traducción propia]" (Mertsching 2018, 4).

Ante esta falta de regulación vinculante a nivel comunitario, fue Alemania el primer país miembro de la Unión Europea en promulgar una ley vinculante que aborde específicamente el discurso de odio en línea. Así la *Netzwerkdurchsetzungsgesetz* (cuya traducción es "Ley de aplicación de la red" y, en adelante nos referiremos como "Ley Alemana"), entró en vigor en octubre del 2017 y tiene como objetivo "luchar de forma más eficaz contra los delitos motivados por el odio, las noticias falsas punibles penalmente y otros contenidos ilícitos en las redes sociales [traducción propia]" (Ministerio Federal de Justicia y Protección al Consumidor Alemán 2020) . En cuanto a su alcance, en la Sección 1 la Ley Alemana explica que se aplicará a la redes sociales que tengan más de dos millones de usuarios registrados en Alemania, consideradas de esta manera como líderes de opinión y además protegiendo así a plataformas pequeñas de los altos costos de implementación de las obligaciones impuestas a las demás plataformas. Un aspecto interesante es que dicha sección carece de una definición de qué considera discurso de odio ilegal y, en vez, para establecer la ilegalidad del contenido remite a ciertos delitos existentes en su Código Penal.

Al igual que el Código, la Ley Alemana contiene una lista de disposiciones a cumplir por los proveedores de redes sociales, en este caso vinculantes. Así, la Sección 3 exige a los proveedores de redes sociales a mantener un procedimiento efectivo y transparente para presentar las quejas sobre contenido ilegal y si el proveedor comprueba que dicho contenido denunciado es manifiestamente ilegal deberá eliminar (en cuyo caso debe conservar el contenido como prueba) o bloquear el acceso a este dentro de las 24 hs posteriores a la recepción de la denuncia. Asimismo, las redes sociales también deben remitir la decisión sobre el contenido ilegal a una institución de autorregulación y luego aceptar la decisión de la última. Por su parte, ante cualquier decisión tomada sobre el contenido denunciado, debe notificar tanto al denunciante como al usuario sobre su decisión y los motivos de esta. Por último, a diferencia

del Código de Conducta, esta ley cuenta con un sistema de sanciones según el cual, si las redes sociales incumplen con sus obligaciones, la multa puede ser de hasta 50 millones de euros. [36]

En suma, la Ley Alemana configura un claro ejemplo de una ley pública que apunta abordar la lucha contra el discurso de odio en línea sometiendo a las redes sociales a diferentes obligaciones bajo pena de multa. Como analizaremos más adelante, normas de este estilo son muy controversiales y despiertan la preocupación por varias razones. Es importante, a la hora de tomar esta ley como ejemplo, tener especialmente en consideración la historia alemana y la consecuente legitimidad que adquiere la regulación del discurso del odio. Así refiere Rosenfeld, quien explica que el enfoque alemán en la lucha contra el *hate speech* es el producto, por un lado de la concepción de su Constitución la cual circunscribe a la libertad de expresión frente a otros valores como la dignidad humana y, por otro lado, del deseo de prevenir el resurgimiento de la historia del Tercer Reich y su virulenta propaganda de odio que culminó en el genocidio Nazi (Rosenfeld 2003).

2.1.2 Reflexiones sobre la regulación pública del discurso del odio en las redes sociales

Luego de haber analizado los diversos enfoques que se le puede dar a la regulación de las plataformas como intermediarios por parte de los poderes públicos para lidiar con el problema del *hate speech* en Internet, nos adentraremos en analizar cuáles pueden ser los los riesgos y benevolencias de que sean los Estados (a nivel estatal o de manera conjunta a través de instituciones internacionales) quienes lleven a cabo el control del contenido *online* a través de la imposición de sanciones a las plataformas a fin de limitar la proliferación del discurso del odio.

[36] Toda la información sobre la Ley Alemana fue obtenida de la página oficial del Ministerio Federal de Justicia y Protección al Consumidor. https://www.bmjv.de/DE/Themen/FokusThemen/NetzDG/NetzDG_EN_node.html

Para comenzar, hay ciertos argumentos que nos alertan de los riesgos de que sean los Estados quienes regulen el contenido compartido a través de las redes sociales. Como es de esperar, el mayor problema frente a la regulación estatal es el riesgo de la censura política a la libertad de expresión. Las redes sociales configuran un ámbito de discusión fundamental dado que las mismas están exentas de censura estatal, cada uno es libre de subir el contenido que desea, teniendo como única restricción las propias políticas de la red social, reglas a las que uno acuerda en el momento en que decide acceder a esta. Así explica Vuarambon, "en el ámbito político, Facebook alcanzó su clímax con la llamada primavera árabe, que comenzó en 2011 con la primera manifestación contra el régimen de Mubarak (Egipto) convocada a través de la red social. De ahí en adelante, en países como Venezuela, donde la libertad de prensa está restringida, las redes sociales pasaron a ser el ámbito de discusión y de denuncias contra las limitaciones de libertades. Este fenómeno permitió conocer la realidad de muchos países cuyos regímenes están manchados por la corrupción, la censura política, los ataques a la libertad de expresión, la violación de derechos humanos, entre otras cosas" (Vuarambon, s.f.). Por esta razón, uno de los temores fundamentales a la hora de pensar la regulación estatal de los contenidos subidos a las redes sociales es la probabilidad de que esto conlleve a abusos políticos en donde, tras el ropaje del discurso del odio, se termine justificando la censura de ciertas expresiones disidentes al gobierno de turno o al discurso de los grupos dominantes[37]. Por su parte, si los Estados deciden llevar a cabo la regulación de las redes sociales a escala internacional, mediante una convención u otro instrumento que sirva a tales fines, el fenómeno de la censura se puede propagar a escala mundial. Es por esto que Balkin explica que "actualmente, Internet se rige principalmente por los valores del régimen menos censurable, el

[37] Para el diputado español, Manuel Mariscal, "la expresión "discursos de odio" es un invento del consenso progre para perseguir a los que no piensan como ellos en el único lugar que aún no controlan: las redes sociales" (@MariscalZabala, 26 de octubre de 2020).

de Estados Unidos. Si los estados nacionales pueden hacer cumplir el filtrado, el bloqueo y la desvinculación global, Internet eventualmente será gobernado por el régimen más censurador. Esto socavará el bien público global de una Internet libre [traducción propia]" (Balkin 2018, 1206).

Asimismo, quienes defienden esta postura también justifican el peligro de delegar en manos del poder estatal la determinación de qué expresiones u opiniones están permitidas y cuáles prohibidas respaldándose en "la verificación histórica de los errores cometidos en el pasado –en épocas en que se persiguieron o censuran ciertas ideas religiosas o políticas (...), que más tarde se consideraron valiosas o importantes" (Marciani Burgos 2013, 165). Así, resulta preferible un discurso amplio, libre y tolerante hacia la diversidad de ideas dado que, en definitiva, expresiones que pueden llegar a resultar repudiables, como por ejemplo expresiones racistas, son también parte del discurso público en una democracia. Por eso, en aras de proteger la libertad de expresión, debemos dejar en manos de la deliberación pública que expresiones u opiniones resultan repudiables, como puede ser el *hate speech*, siendo el mismo mercado de ideas[38] el que las rechace. En esta línea, Boix Palop (2016) explica que hay que aceptar y tolerar la difusión de todo tipo de expresiones a través de las redes sociales antes que establecer una censura estatal sobre ciertas ideas u opiniones, expresando que el hecho de que dichas plataformas "faciliten ahora que esa misma expresión pueda llegar mucho más lejos y alcanzar a muchas más personas no debiera ser una mala noticia en sí misma, sino más bien una buena, en la medida en que se incrementan las posibilidades de que sea conocida y rebatida por un mayor número de personas [...] contribuyendo así a mejorar el debate público y la construcción de la opinión pública libre en una sociedad democrática" (Boix Palop 2016, 71). Este riesgo de la regulación Estatal se puede ver ilustrado en el caso de Jörg Rupp, un activista

[38] *Ibid.*, 6

político que, por aplicación de la *Netzwerkdurchsetzungsgesetz*, la Ley Alemana, su cuenta de Twitter fue inhabilitada tras un tweet irónico en donde utilizaba el lenguaje de los grupos de derecha con el objetivo de demostrar su crueldad. [39]

Por último, quienes critican que sean los Estados quienes regulen los contenidos que circulan en las redes sociales, argumentan que tratándose de un derecho fundamental, debe ser un juez quien decida qué se considera una expresión de odio capaz de constituir un límite a la libertad de expresión y que no sea el derecho administrativo quien invada dichas competencias judiciales (Almeida, 2020). De esta manera, a través de una ley que imponga obligaciones a las redes sociales respecto a contenidos generados por terceros, se le estaría solicitando a prestadores privados que decidan que configura un contenido ilícito. Aún peor, mediante leyes que presionen a las compañías a suprimir rápidamente contenidos que constituyan *hate speech* bajo amenaza de ser gravemente sancionadas en caso de no cumplir, en definitiva, sería el propio gobierno quien decide el valor de los contenidos. Así, explica Carrasco que el principal problema es que "se trata de una ponderación complicada de derechos, que es la razón para que dirima estas cuestiones un juez. Un prestador, en caso de duda, puede acabar retirando contenidos para no llegar a ser responsable, lo cual al final supone una restricción de derechos fundamentales" (Carrasco 2020). En definitiva, quienes proponen esto, argumentan que en lo que hay que enfocarnos es en reforzar el poder judicial de los Estados para que puedan tratar los temas de *hate speech* a través de las redes sociales con la urgencia que requieren.

Por su parte, entre quienes defienden la necesidad de leyes estatales que regulen a las redes sociales en cuanto al contenido que difunden en sus plataformas, uno de los argumentos se basa en que dada la importancia de la cuestión, la regulación no puede quedar librada a la

[39] Información obtenida de. https://www.clarin.com/new-york-times-international-weekly/vigilancia-online-europa-inquieta-regular-internet-equivale-censurarla_0_6jO_e887W.html.

propia red social, si no que es necesario que el Estado a través del Poder Legislativo establezcan un marco de regulación del discurso de odio en Internet. Así, Angela Merkel, canciller alemana, expresó - a través de su portavoz Steffen Seibert- que las compañías proveedoras "tienen la gran responsabilidad de garantizar que la comunicación política no se vea envenenada por el odio, la mentira o la incitación a la violencia [pero esto] solo puede tener lugar según la ley y en el marco definido por el legislador, y no según la decisión de la dirección de las plataformas de medios sociales" (Seibert 2021)[40]. Por su parte, Bruno Le Maire, actual Ministro de Finanzas de Francia, expresó en esta línea que "la regulación de los gigantes digitales no puede hacerse por la misma oligarquía digital" (Le Maire 2021)[41].

Otra de las preocupaciones que expresan quienes proclaman la necesidad de regular a nivel estatal el *hate speech* en las redes sociales, basan sus argumentos en las graves consecuencias que trae aparejado el discurso de odio sobre sus víctimas y en los eventos que puede desencadenar en el mundo fuera de línea. Teniendo en cuenta esto, quienes pregonan esta postura sostienen que se trata de una cuestión tan delicada que la responsabilidad de regular debe recaer sobre el Estado y no sobre las compañías privadas. Por eso es que Rosenfeld (2003) critica el enfoque estadounidense al subestimar el daño potencial que tiene el discurso de odio o sobreestimar el potencial de la deliberación racional[42]. Una de las principales consideraciones

[40] Para más información, ver: https://elpais.com/internacional/2021-01-11/merkel-ve-problematica-la-suspension-de-las-cuentas-de-trump-en-redes-sociales.html

[41] Para más información, ver: https://www.france24.com/es/minuto-a-minuto/20210113-un-twitter-sin-trump-aviva-el-debate-sobre-la-regulaci%C3%B3n-de-los-gigantes-de-la-red

[42] En esta línea, Rosenfeld agrega que "en términos de impacto, dada su larga historia de tensiones raciales, es sorprendente que Estados Unidos no muestre mayor preocupación por los daños a la seguridad, dignidad, autonomía y bienestar que el discurso de odio oficialmente tolerado causa a su minoría negra. Del mismo modo, el enfoque del discurso de odio en Estados Unidos parece descartar indebidamente el impacto pernicioso que el discurso de odio racista puede tener en los sentimientos racistas latentes o latentes que aún albergan un segmento no despreciable de la población blanca [traducción propia]' (Rosenfeld 2003, 1559).

que tiene en cuenta el autor a la hora de criticar la postura estadounidense y defender la necesidad de una mayor regulación del discurso de odio, es el hecho de que las democracias constitucionales son cada vez más diversas (en términos raciales, culturales, religiosos, lingüísticos, etc.) lo cual hace que la cohesión social sea más precaria, contexto propicio para exacerbar los potenciales males del *hate speech.* Esto torna especialmente relevante, según el autor, cuando en las sociedades pluralistas la "tiranía de la mayoría" logra dominar el discurso silenciando el de las minorías y amenazando así su libertad de expresión[43]. En esta línea, el autor explica que por lo general el discurso de odio tiende a tener vínculos estrechos con las opiniones de las mayorías y, por ende, el dominio de su discurso "representa la mayor amenaza para la autoexpresión desinhibida y el debate político sin restricciones en una política pluralista contemporánea [traducción propia]" (Rosenfeld 2003, 1561). Por esta razón, sostiene que en las sociedades pluralistas contemporáneas donde el discurso del odio puede extenderse a escala mundial a través de las redes sociales, es necesario que el Estado intervenga y lleve adelante una lucha activa contra estas expresiones mediante su regulación. Así, expresa que "el Estado ya no puede justificar el compromiso con la neutralidad, sino que debe abrazar el pluralismo, garantizar la autonomía y la dignidad y esforzarse por mantener un mínimo de respeto mutuo [traducción propia]" (Rosenfeld 2003, 1566-1567).

[43] Este enfoque en el cual cobra relevancia para la libertad de expresión tener en cuenta el efecto silenciador del discurso de ciertos grupos sobre otros, fue especialmente desarrollado por el profesor Owen Fiss. El mismo plantea a la libertad de expresión como un derecho público cuyo fin es asegurar que el debate público sea lo suficientemente rico en el que todas las voces y puntos de vista puedan ser escuchados. En este sentido, el autor sostiene que dado que con "el discurso de odio […] el temor es que el discurso hará imposible que estos grupos desfavorecidos participen en la discusión [traducción propia]" (Fiss, 1996, p.16), entonces, la regulación del mismo irónicamente ampliará en lugar de reducir la libertad de expresión dado que se favorece al debate público.

2.2 La regulación privada del *hate speech* en las redes sociales

Más allá de la conveniencia o no de regulación estatal del discurso del odio en las redes sociales, la realidad es que hoy en día lo que sí está sucediendo es que son las propias compañías las que establecen las políticas respecto de qué tipo de contenido aceptan en sus plataformas. En este sentido, durante años, los grandes de Silicon Valley han adoptado reglas y políticas de contenido que reflejaban los valores estadounidenses a favor de la libertad de expresión. Si bien las compañías proveedoras de redes sociales admitieron desde un principio ciertas excepciones- relacionadas con la pornografía no consensuada o el acoso *online*-, la realidad es que estas han sido muy limitadas, manteniendo las compañías una arraigada cultura corporativa ampliamente comprometida con la libertad de expresión (Citron 2018).

Sin embargo, en los últimos tiempos y en especial como consecuencia de su acuerdo con la Comisión Europea, las compañías se han apartado de esta postura dejando atrás sus pretensiones de ser plataformas neutrales, para adoptar ciertas políticas que regulan el contenido de las publicaciones generadas por sus usuarios. En este sentido, ilustrando este cambio en cuanto a la tolerancia frente a los contenidos, Karen White- jefa de políticas públicas de Twitter- expresó que "la conducta de odio no tiene cabida en Twitter [...] existe una clara distinción entre la libertad de expresión y la conducta que incita a la violencia y al odio [...] que infringen las Reglas de Twitter" (White 2016) [44] . Por su parte, la directora de Gestión de Políticas Globales de Facebook, Monika Bickert, en esta misma línea explicó: "como dejamos

[44] De hecho, en la política relativa al discurso del odio de Twitter se indica: "Nuestro compromiso es combatir el abuso motivado por el odio, el prejuicio o la intolerancia, en particular, el abuso cuyo objetivo es silenciar las voces de quienes han sido históricamente marginados. Por esta razón, prohibimos el comportamiento abusivo dirigido hacia las personas con base en las categorías protegidas. Si ves contenido en Twitter que creas que incumple nuestra política relativa a las conductas de incitación al odio, denuncialo". Para más información sobre la polícia contra el *hate speech* de la plataforma, visitar el siguiente link: https://help.twitter.com/es/rules-and-policies/hateful-conduct-policy

claro en nuestras Normas de la comunidad, no hay lugar para el discurso de odio en Facebook. Instamos a las personas a que utilicen nuestras herramientas de denuncia si encuentran contenido que creen que infringe nuestras normas para que podamos investigar [...] y toma[r] medidas rápidas" (Bickert 2016)[45] [46]. Un ejemplo controversial de las amplias restricciones que Twitter ha estado implementando en este sentido fue el bloqueo de la cuenta de Donald Trump, entonces Presidente de los Estados Unidos.

Cuando hablamos de regulación privada por parte de las redes sociales, en algún sentido no cabe hacer referencia al concepto de censura relacionado con el derecho de la libertad de expresión ya que en definitiva se trata de una compañía privada la cual diseña sus propios términos de servicio y está en la decisión de los usuarios adherirse a la misma o no.[47] De todas formas, tal como remarca Kate Ruane -asesora legislativa estadounidense-, dado el carácter monopólica de las empresas proveedoras y el potencial de la misma para permitir la expresión de millones de voces, debería alarmarnos cuando estas compañías tienen el poder de decidir qué contenido retirar de sus plataformas las cuales se han vuelto indispensables para la expresión de la gente (Ruane 2021)[48].

[45] En la política de Facebook respecto al discurso que incita al odio, se explica: "Creemos que las personas se expresan y se conectan entre sí con mayor libertad cuando no se sienten atacadas por quiénes son. Es por eso que no permitimos el lenguaje que incita al odio en Facebook, ya que crea un entorno intimidatorio y excluyente que, en algunos casos, puede incitar a la violencia en la vida real". Para más información sobre la política contra el discurso de odio de la plataforma, visitar el link a continuación: https://www.facebook.com/communitystandards/hate_speech

[46] Ambos comentarios fueron publicados en el comunicado de prensa del 31 de mayo de 2016 de la Comisión Europea. El mismo se puede encontrar en el siguiente link: https://ec.europa.eu/commission/presscorner/detail/en/IP_16_1937

[47] *Ibid.*, 21.

[48] Citado en "Trump, Twitter y el gran debate sobre la "censura": quién tiene el poder para marcar qué puede decirse en las redes sociales". *Xataka*, 18 Enero 2021. https://www.xataka.com/legislacion-y-derechos/trump-twitter-gran-debate-censura-quien-tiene-poder-para-marcar-que-puede-decirse-redes-sociales

Uno de los episodios que han generado mayor controversia en este respecto fue que una empresa privada como Twitter pudiera silenciar al entonces presidente de los Estados Unidos, Donald Trump al bloquear su cuenta en razón de presuntos tuits que incitaron a violencia en el marco de la toma del capitolio[49]. En este sentido, Jack Dorsey- CEO de Twitter- a través de una serie de tuits, en primer lugar justificó el accionar de la compañía en la protección de la seguridad pública, expresando que "el daño fuera de línea como resultado del discurso en línea es demostrablemente real, y lo que impulsa [su] política y cumplimiento sobre todo" (@jack 13 de enero de 2021). En esta línea, Dorsey expresó mediante la plataforma que si bien estas acciones fragmentan la conversación pública, limitan su potencial y sientan un peligroso antecedente al darle el poder a una empresa sobre la conversación pública a escala mundial, de todas maneras Twitter no deja de ser solo una parte de dicha conversación la cual se lleva a cabo a través de Internet y en caso de no estar de acuerdo con sus políticas de contenido, los usuarios pueden expresarse en otra plataforma de *online.*

Como bien se mencionó anteriormente, más allá de lo preocupante que nos pueda resultar el poder que tiene una red social dirigida por una compañía privada para regular el discurso *online*, la realidad es que al ser entes privados estos pueden regular el contenido difundido a través de sus plataformas mediante sus términos de servicio sin poner en riesgo la libertad de expresión ya que el fundamento de este derecho radica en proteger al individuo frente al poder de censura estatal.[50] Sin embargo, el rol de las redes sociales como reguladores

[49] El ataque al Capitolio de los Estados Unidos ocurrió el 6 de enero de 2021 cuando partidarios del ex presidente estadounidense, Donald Trump, irrumpieron en el capitolio ocupando durante horas el edificio . Para más información: https://www.nytimes.com/spotlight/us-capitol-riots-investigations

[50] *Ibid.*, 22. Tal como se explicó previamente, este trabajo asume la postura estadounidense según la cual los derechos constitucionales, como lo es la libertad de expresión, protegen a los individuos frente al poder del Estado y no frente a otros individuos privados.

Ahora bien, resulta importante reconocer que más allá del derecho formal, a efectos prácticos las políticas de contenido de las redes social sí pueden afectar la libertad de expresarse de las personas. Y, si bien en Internet hay inmensas posibilidades para poder comunicar ideas, la realidad es que

de la expresión a través de Internet torna especialmente relevante al reconocer que hoy en día ya no se puede hablar lisa y llanamente de regulación privada. Este punto lo ilustra Germán Turuel-profesor de derecho constitucional- al expresar que uno de los riesgos hoy en día con el pluralismo en internet es el riesgo de censura privada. Aún más, y esto preocupa, cuando esa censura privada viene impulsada o mediatizada por los poderes públicos" (Teruel 2020).

De esta manera, el principal riesgo a la hora de delegar el poder de la regulación de los contenidos *online* a las plataformas privadas es que hoy en día la realidad es que dicho control no se lleva a cabo libremente por la *tech companies* sino que, por el contrario, es el resultado de la presión llevada a cabo por parte de los poderes públicos. Así, explica Citron que dado el reciente aumento de los ataques terroristas y de grupos de odio, los reguladores de la Unión Europea han empezado a presionar cada vez más a las compañías para que ajusten sus políticas de contenido y asuman un rol más activo en la tarea limitar el material extremista y las expresiones de odio en sus plataformas. Por ende, los cambios en la manera de confrontar el discurso de odio llevados a cabo por las compañías tecnológicas en los últimos años no fueron el resultado de elecciones voluntarias e independientes en respuesta a demandas del mercado, sino que más bien, fueron el resultado de la coerción de los poderes públicos mediante la amenaza de imponerles regulación estatal (Citron, 2018). Ejemplo de ello es el anuncio de Jourova quien, en el marco de una evaluación llevada a cabo por la Comisión Europea en 2016

plataformas como Facebook, Twitter e Instagram tienen un peso excepcional para la difusión de ideas y la posibilidad de participar en el debate público.

Esto se puede ver ilustrado en el caso del blog creado por el ex presidente estadounidense denominado "*From the Desk of Donald J. Trump"* (cuya traducción es "Desde el Escritorio de Donald J. Trump"). Este fue creado por Trump luego de que sus cuentas de Twitter y Facebook hayan sido bloqueadas con el fin de poder seguir compartiendo sus pensamientos e ideas a sus partidarios. Sin embargo, el sitio fue rápidamente suspendido por el mínimo alcance de lectores que obtuvo. Para más información al respecto, ingresar a: https://www.nytimes.com/2021/06/02/us/politics/trump-shuts-down-blog.html

en cuanto al manejo de la incitación al odio por parte de las empresas de redes sociales, advirtió que si las compañías "quieren convencer[la] a [ella] y a los ministros de que el enfoque no legislativo puede funcionar, tendrán que actuar con rapidez y hacer un gran esfuerzo en los próximos meses" (Clark, 2016).

Como podemos notar, las concesiones llevadas a cabo por los gigantes de Silicon Valley a los reguladores europeos ponen en riesgo a la libertad de expresión de los individuos ya que, en última instancia, los poderes públicos delegan en las entidades privadas para que lleven a cabo actos de censura en su nombre (Coche 2018). Además, un factor que complica todavía más este panorama es el hecho de que, al tratarse de plataformas *online*, la censura impulsada por los estados a través de los términos de servicio de las compañías privadas adquiere propiedades que la pueden convertir en una censura a nivel global. Así explica Citron que "al insistir en cambios en las reglas y prácticas de expresión de las plataformas, los reguladores de la UE han ejercido su voluntad en todo el mundo. A diferencia de las leyes nacionales que se aplican sólo dentro de las fronteras de un país, los términos de servicio se aplican dondequiera que se acceda a las plataformas [...] tiene el potencial de resultar en censura mundial [traducción propia]" (Citron 2018, 1038).

En definitiva, podemos concluir que hoy en día resultaría un tanto ingenuo referirnos a la posibilidad de pensar en que, en caso de creer que la regulación por parte de los poderes públicos es peligrosa para la libertad de expresión, la otra opción es dejar que la cuestión se regule por las fuerzas del mercado y en base a las decisiones de las compañías privadas. Por el contrario, como vimos a través de este apartado, la realidad es que en la actualidad los cambios de política de contenido llevados a cambio por las redes sociales no son el producto de una decisión voluntaria, sino que detrás del ropaje de lo privado se esconden presiones estatales que ponen en riesgo la libertad de expresión. Mediante dichas presiones de los poderes

públicos para que las compañías privadas adopten las preferencias gubernamentales en la materia "los actores estatales disfrutan de las ventajas del poder gubernamental al tiempo que evitan el desorden de los debates políticos y las audiencias judiciales [traducción propia]" (Citron 2018, 1061).

3. Reflexiones y alternativas para combatir el discurso de odio *online* a través de la regulación pública

A lo largo de este trabajo se recorrió el estado de la cuestión actual en cuanto al discurso de odio en las redes sociales, las consecuencias a las que este puede conllevar si adquiere la escala masiva que las redes sociales le posibilitan y frente a esto los principales paradigmas a la hora de regularlo. Una vez que tenemos en cuenta las graves consecuencias que puede generar el discurso de odio difundido por las redes sociales podemos argumentar que dichos riesgos que conlleva en el mundo *offline* justifican la necesidad de reflexionar sobre la posible necesidad de cierto tipo de regulación que no puede quedar totalmente en manos de privados.

Por su parte, como se enfatizó, a la hora de pensar en una posible regulación diseñada por un ente público, el principal factor que genera preocupación es que la misma se convierta en un instrumento de censura estatal que vulnere el derecho fundamental de la libertad de expresión. Sin embargo, el otro extremo que defiende el ciber libertarismo que lucha por la libertad en el espacio de Internet, hoy en día ya no parece posible dado que, como vimos precedentemente, las compañías proveedoras ya no diseñan sus políticas de contenido de manera libre e independiente, sino que estas son el resultado de presiones políticas, especialmente provenientes de la Unión Europea. Así, luego de lo analizado es difícil pensar en la posibilidad de una regulación puramente privada por parte de las compañías de redes sociales, sin que en el fondo en definitiva se estén encubriendo presiones estatales, lo cual exacerba aún más los riesgos implicados con la regulación pública del tema. Es por ello que

considero que, en el estado de la cuestión actual, pensar en una alternativa de regulación pública deviene necesario. De tal manera, en lo que quede de este trabajo, plantearemos y analizaremos ciertos estándares básicos a tener en cuenta a la hora de pensar en el diseño de una posible regulación pública que le haga frente al problema del *hate speech* en las redes sociales, pero a su vez protegiendo el derecho fundamental de la libertad de expresión.

4. Rumbo a un posible diseño de regulación pública del discurso del odio en las redes sociales: alternativas y formas de mitigar sus riesgos

4.1 El régimen de responsabilidad de los intermediarios

Las redes sociales ostentan un rol de intermediarios con un papel fundamental para la difusión de mensajes de odio. Estos se caracterizan por permitir a las personas conectarse a Internet y publicar su contenido sin ser ellos precisamente los productores de dicho contenido. Así, explica Wenguang que "existen muchos tipos diferentes de intermediarios, incluidos los proveedores de acceso a Internet, los proveedores de alojamiento web, las plataformas de redes sociales y los motores de búsqueda [traducción propia]" (Wenguang 2018, 347).

Dado este rol primordial de los intermediarios, como punto de partida para reflexionar acerca del desarrollo de un modelo de regulación para combatir el discurso de odio en línea, debemos analizar cuál sería el régimen de responsabilidad de los intermediarios más adecuado. En este sentido, cabe reflexionar acerca de si es posible atribuirle responsabilidad a los intermediarios frente al contenido generado por terceros de la cual estos sirven como plataforma para su difusión, y en caso de ser así, qué tipo de responsabilidad sería apropiado atribuirles.

Existen al menos tres modelos que dan una posible respuesta a la disyuntiva acerca de la responsabilidad de los intermediarios por el contenido creado por terceros. Por un lado, la

postura a favor de eximir de responsabilidad a las redes sociales como intermediarios es cada vez más criticada y se reputa como no sostenible hoy en día (Wenguang 2018). Entre las críticas a esta postura, Wenguang (2018) explica que en la era de la web 3.0 la teoría de la inmunidad ha quedado desactualizada porque la misma se basaba en argumentar que los intermediarios cumplían un rol neutral en cuanto al contenido generado por terceros del cual servían como plataforma para su difusión. En este sentido, este modelo interpreta a los intermediarios como carentes del control acerca del contenido generado por terceros dado que eran incapaces de verificar la legalidad sobre todo el contenido publicado por sus usuarios. Sin embargo, tal como explica el autor " en la era de la web 3.0, muchos intermediarios hacen mucho más que servir como meros conductos pasivos en la forma en que diseñan sus aplicaciones para recopilar, analizar y clasificar los datos de los usuarios por sus propias razones comerciales. Hasta cierto punto, dan forma y gestionan activamente el contenido y el comportamiento de los usuarios [traducción propia]" (Wenguang 2018, 350). Asimismo, explica que con el avance y desarrollo tecnológico, hoy en día las redes sociales como intermediarias cuentan con los recursos necesarios para poder monitorear, filtrar y eliminar el contenido ilegal de sus plataformas.

Incluso en los Estados Unidos esta doctrina plasmada en la sección 230, fue puesta en cuestionamiento en el último tiempo[51]. Así, en octubre de 2020 se citó a una audiencia frente al Senado Estadounidense a los principales directivos de Google, Facebook y Twitter en donde tanto la entonces administración de Donald Trump como el actual presidente demócrata, Joe Biden, llevaron a cabo esfuerzos para atacar la sección 230, en base a diferentes argumentos.

[51] En este sentido, explica Wenguang que " la doctrina original de inmunidad [...]ahora enfrenta grandes desafíos y debates sobre si todavía se adapta al mundo de hoy. Hay una reforma en curso de la responsabilidad de los intermediarios que se debe observar. Por ejemplo, los tribunales de EE. UU. han comenzado a apartarse de su lectura amplia de la Sección 230 y a prestar más atención a las formas en que los intermediarios en línea "diseñan" a los usuarios el contenido y las transacciones en línea a través de algoritmos. Los intermediarios perderían su inmunidad si son "responsables, total o parcialmente, de la creación o desarrollo de información [traducción propia]" (Wenguang 2018, 349).

Así, por un lado, “el ala demócrata cree que la Sección 230 es demasiado indulgente y permite la proliferación de contenido abusivo e incitaciones a la violencia. Los republicanos, por otro lado, sostienen que se usa de una manera injusta y que da a las plataformas el derecho de reprimir las publicaciones más conservadoras, sobre todo las del presidente Trump” (Vega 2020).

Por otro lado, se puede pensar en un tipo de responsabilidad estricta y objetiva, a partir de la cual mediante la regulación se prevea un deber de monitoreo y vigilancia general por parte de las redes sociales a las cuales se les imponga algún tipo de sanción por la no eliminación de contenidos considerados ilegales publicados por sus usuarios. Esta es la línea actual de la Unión Europea según Thierry Breton, quien develó los planes que tiene la Comisión Europea de multar a las redes sociales por los contenidos considerados ilegales a través de una propuesta de nueva legislación. Así, explica Sabán que “lo más relevante, enmarcado en el contexto de la nueva Ley de Servicio Digitales, Digital Services Act o DSA, es que desde Bruselas quieren hacer que las plataformas de Internet dejen de ser intermediarias para convertirse en responsables del contenido que los usuarios publicamos (...) en cuanto a su cumplimiento (...) serán extremadamente estrictos" (Sabán 2020). Por su parte, en el artículo se explica que a través de dicha directiva lo que se busca es establecer una mayor responsabilidad a los intermediarios para de esta manera fomentar su proactividad a la hora de retirar contenidos de sus plataformas[52].

[52] Para poder reflexionar acerca de un tipo de responsabilidad objetiva y las consecuencias que esto podría acarrear podemos hacer una comparación con la Copyright Directive, una propuesta de directiva de la Unión Europea para proteger los derechos de autor en Internet, la cual fue aprobada en 2018 y luego nuevamente en 2019. En el artículo 17 de la susodicha (ex art.13) se indica que los prestadores de servicios para compartir contenidos en línea serán responsables de los actos no autorizados de comunicación al público a menos que demuestren que han hecho “los mayores esfuerzos por garantizar la indisponibilidad de obras y otras prestaciones específicas respecto de las cuales los titulares de derechos les hayan facilitado la información pertinente y necesaria”.

Una regulación que haga recaer este tipo de responsabilidad tan estricta sobre las redes sociales tiene grandes probabilidades de poner en riesgo la libertad de expresión y la libre interacción entre sus usuarios dado que puede generar que dichas plataformas terminen imponiendo filtros más estrictos sobre los contenidos generados por sus usuarios con el fin de evitar incurrir en responsabilidad. En este sentido, tal como explica Wenguang (2018) las críticas sobre las responsabilidades tan estrictas sobre los intermediarios y las obligaciones de monitoreo se basan en el hecho de que en última instancia esto podría conducir a la censura privada por parte de las compañías, con su inherente riesgo en cuanto a la libertad de expresión. Así, el autor explica que "de hecho, los intermediarios hoy en día son capaces de monitorear y controlar el acceso al contenido en línea a través de tecnología de filtrado o bloqueo. Sin embargo, la responsabilidad desproporcionada o las obligaciones de vigilancia podrían llevar a los intermediarios a bloquear incluso el contenido legal y, por lo tanto, causar un riesgo inherente de restringir la libertad de expresión sin la transparencia o la rendición de cuentas adecuadas [traducción propia]" (Wenguang 2018, 345)[53]. Otra preocupación que versa entre los autores es el hecho de que, en términos comerciales, una regulación tan estricta sobre los intermediarios afectaría a los medianos y pequeños proveedores de redes sociales quienes no podrán competir contra los grandes proveedores a la hora de implementar mecanismos para el

Los peligros de que los intermediarios deban soportar con este tipo de responsabilidad fueron alegados por Polonia en el caso frente al Tribunal de Justicia de la Unión Europea, *Republic of Poland v European Parliament and Council of the European Union* (2019), en el cual alegó que el art. 17 hace necesario "que los proveedores de servicios, para evitar responsabilidades, realicen una verificación automática previa (filtrado) de los contenidos subidos online por los usuarios, por lo que hacen necesaria la implantación de mecanismos de control preventivo. Dichos mecanismos atentan contra la esencia del derecho a la libertad de expresión e información [traducción propia]" (2019).

[53] La autora agrega que "a los defensores de la doctrina de la inmunidad de intermediarios les preocupa que imponer responsabilidades a los intermediarios por no monitorear o censurar el contenido ilegal generado por el usuario podría tener un efecto paralizador en la libertad de expresión, ya que los intermediarios tienden a pecar de cautelosos y eliminar o bloquear incluso algún contenido legal para protegerse y evitar demandas. Esta fue también una de las razones para la promulgación de la Sección 230 CDA y la amplia lectura de esta disposición legal por parte de los tribunales [traducción propia]" (Wenguang 2018, 352).

monitoreo de los contenidos subidos en sus plataformas, lo cual en última instancia podría inhibir la tecnología e innovación (Wenguang 2018)[54].

Una tercera opción de responsabilidad de los intermediarios es aquella que podemos llamar como "responsabilidad subjetiva". En base a la misma, se protegería a las redes sociales como intermediarios de ser responsables ante el contenido publicado por terceros solo en caso de que las mismas actúen de manera diligente frente al contenido de odio. En este sentido, no cabría imponerles una obligación general de monitoreo proactivo a las redes sociales por el contenido ilegal difundido por sus usuarios, siempre que las mismas actúen simplemente como conductores pasivos de dicho contenido. Por su parte, las redes sociales como intermediarios podrían perder esta inmunidad de no cumplir y actuar en base a ciertos criterios. Uno de estos podría ser que las mismas sean responsables si no eliminan o inhabilitan el acceso a información ilegal cuando estas obtengan conocimiento de la misma[55]. Otros criterios podrían ser, adoptar un sistema de notificación y mecanismos de monitoreo eficientes, tener personal capacitado respecto al *hate speech*, entre otros. Así, explica Wenguang que "este modelo busca encontrar un término medio que, por un lado, reconozca los beneficios de la protección de

[54] Sobre esto, Dorsey expresó su preocupación al afirmar que erosionar las bases de la Sección 230 podría hacer colapsar las comunicaciones por Internet al dejar en pie solo a un pequeño número de grandes compañías tecnológicas y bien financiadas. Para más información al respecto, visitar el siguiente link: **https://www.clarin.com/tecnologia/twitter-google-facebook-despegaron-contenido-publican-usuarios-retados-congreso_0_Bg8-BL7Sq.html**

[55] En el *leading case* Rodríguez, María Belén c/ Google Inc. y otro s/ daños y perjuicios, en el cual la actora demandó a Google por el uso comercial no autorizado de su imagen en relación a páginas de contenido erótico, nuestra Corte Suprema explicó que "la libertad de expresión seria mellada de admitirse una responsabilidad objetiva que -por definición- prescinde de toda idea de culpa y, consiguientemente, de juicio de reproche a aquél a quien se endilga responsabilidad [...] Que sentado lo expuesto, hay casos en que el "buscador" puede llegar a responder por un contenido que le es ajeno: eso sucederá cuando haya tomado efectivo conocimiento de la ilicitud de ese contenido, si tal conocimiento no fue seguido de un actuar diligente [...] A partir del momento del efectivo conocimiento del contenido ilícito de una página web, la "ajenidad" del buscador desaparece y, de no procurar el bloqueo del resultado, sería responsable por culpa" (2014, considerandos 16 y 17).

responsabilidad y, por otro lado, defina ciertos roles para los intermediarios con respecto al contenido ilegal [traducción propia]" (Wenguang 2018, 347-348).

En vistas a estos tres modelos de régimen de responsabilidad, para una posible regulación del *hate speech online* se podría adoptar una sistema de responsabilidad subjetiva en donde se fijen ciertos criterios y condiciones bajo los cuales las redes sociales ejerzan su poder y cumplan con su responsabilidad moderadora del discurso del odio que pueda traer consecuencias en el mundo *offline*, pero sin dar lugar a una supresión de discursos que ponga en peligro la libertad de expresión. Asimismo, al establecer ciertos criterios a cumplir, en vez de una tarea de monitoreo general, se incentivaría a las plataformas a ser diligentes sin imponerles una tarea casi imposible en sus manos. Es por eso que, en los apartados subsiguientes, se analizarán diferentes alternativas y puntos a tener en cuenta a la hora de pensar en una regulación pública que regule y responsabilice a las redes sociales como intermediarios frente a la necesidad de combatir el discurso de odio en línea.

4.2 Regulación local o global

Una de las primeras cuestiones a plantearse a la hora del diseño de una regulación pública que responsabilice a los intermediarios por los contenidos difundidos por terceros es el hecho de si dicha normativa debe aplicarse de manera local, en donde la difusión del contenido odioso se limite acorde a las particularidades idiomáticas, culturales, ideológicas e históricas de cada Estado. O por su parte, que la regulación sea mundial, una regla única que funcione y se aplique de manera general para todos los países y que de esta manera conlleve a prohibiciones de contenido generalizadas.

Por un lado, tal como argumenta Rosenfeld, "dada la tendencia hacia la globalización y el alcance transnacional instantáneo de Internet, un enfoque puramente contextual parecería

insuficiente, si no totalmente inadecuado [traducción propia]" (Rosenfeld 2003, 1524). En este sentido, podemos sostener que tendría escasa efectividad aplicar la regulación restringida a las fronteras nacionales, cuando en definitiva dada la potencial expansión internacional con la que cuentan las redes sociales, el discurso de odio tendrá efectos transnacionales. Por ejemplo, si propaganda NeoNazi es generada por un usuario desde Estados Unidos y alguien lo denuncia a la red social en cuestión en ese mismo país y solo se prohíbe allí, en definitiva, su eliminación en dicho territorio no prevendrá su propagación hacia otros estados donde estos discursos pueden llegar a través de su transmisión por Internet y adonde pueden resultar aún más peligrosos, como es el caso de Alemania. Por eso, dado el carácter "sin fronteras" del que gozan las redes sociales, resultaría necesario que los estándares que se establezcan a partir de una regulación del discurso del odio *online* tengan que ser aplicados por dichas plataformas a nivel internacional, sin importar si en el caso específico, como lo es EE. UU en el ejemplo, a nivel contextual del país ese discurso podría llegar a resultar protegido. Sin embargo, por otro lado, el hecho de que las redes sociales adapten geográficamente sus condiciones de servicio también tiene su relevancia. Así, tal como explica Citron , adaptar los términos de servicio por país o por región permitiría que se eliminen o bloquee ciertos discursos en países con mayor protección a la libertad de expresión, como lo vimos con el paradigma norteamericano, y por otro lado, permitiría que se bloqueen o eliminen los discursos de odio en regiones con normas más restrictivas de la expresión, como lo es el paradigma europeo (Citron 2018).

Teniendo en cuenta estos dos extremos, considero que una forma de abordar esta cuestión podría ser que, a nivel universal, los términos de servicio apliquen cierto conjunto de principios normativos fundamentales en pos de la protección de la libertad de expresión que aplique a todos los países y regiones por igual. En este sentido, se podría mantener un estándar básico de protección del derecho fundamental y a la vez aplicar ciertas restricciones en cuanto a los mensajes del odio por igual en todos los países. Así, por ejemplo, la definición de qué es

considerado *hate speech* podría ser universal. Para definir estas protecciones y límites mínimos, la regulación se podría basar en diferentes convenciones internacionales, las que contemplen las mayores protecciones para la libertad de expresión, pero a su vez teniendo en cuenta otros valores en juego como la protección de la dignidad humana, la igualdad y la autonomía.

Ahora, si bien dado el potencial internacional de Internet resultaría poco eficaz una regulación que se aplique a nivel local en cada país, resulta relevante que a la hora de aplicarla a nivel universal a través de sus términos y condiciones, las redes sociales puedan aplicar las últimas adaptándose a los diferentes contextos sociales, teniendo en cuenta que no todos los discursos del odio implican lo mismo en los diferentes lugares y que sus consecuencias varían dependiendo del entorno[56]. Así, nos explica Rosenfeld que "los estándares de regulación constitucionalmente permisible del discurso del odio deben ajustarse a principios fundamentales que trascienden las diferencias geográficas, culturales e históricas, y al mismo tiempo permanecer lo suficientemente abiertos para acomodar variables históricas y culturales de gran relevancia [traducción propia]" (Rosenfeld 2003, 1565). La importancia de un enfoque de regulación universal del discurso de odio a través de las condiciones de servicio pero que a su vez tenga en cuenta las particularidades de cada región se puede ver ilustrado en el caso de Myanmar que analizamos previamente. En este sentido, tal como lo explica el informe del Consejo de Derechos Humanos de la ONU (2018), en torno al conflicto latente en Myanmar, 6 organizaciones civiles locales expresaron su preocupación por el hecho de que Facebook, la principal red social involucrada en el conflicto, no contaba con suficientes moderadores de contenido que comprendieran el idioma de Myanmar ni tampoco sus matices y contexto dentro del cual se difundieron los mensajes. Así, el informe explica como en junio de 2018 Facebook

[56] Por ejemplo, en el caso Sürek v. Turkey, mencionado previamente, a la hora de evaluar el contenido de odio de las cartas publicadas, la Corte prestó "especial atención a las palabras utilizadas en las cartas y al contexto en el que fueron publicadas. En este último sentido, [tuvo] en cuenta los antecedentes de los casos que se le someten, en particular los problemas vinculados a la prevención del terrorismo [traducción propia]" (1999).

declaró que "había agregado más revisores de idiomas de Myanmar para manejar los informes de los usuarios en todos sus servicios, aumentó el número de personas en la empresa sobre temas relacionados con Myanmar y estableció un equipo especial que trabajara para comprender mejor los desafíos locales específicos y construya las herramientas adecuadas". Por su parte, otra de las variables para que las redes sociales adopten un enfoque contextual a la hora de aplicar las normas relativas al discurso del odio podría ser basarse en las diferentes experiencias históricas. En este sentido, siguiendo con el ejemplo del discurso antisemita, parecería razonable que en la aplicación de prohibición de dichos discursos las plataformas sean más estrictas en su aplicación en Alemania que en Estados Unidos. Así argumenta Rosenfeld explicando que "aunque los judíos estadounidenses y alemanes tienen derecho al mismo grado de dignidad e inclusión dentro de sus respectivas sociedades, se requieren mayores restricciones al antisemitismo en Alemania que en los Estados Unidos [traducción propia]" (Rosenfeld 2003, 1566). El mismo argumento se podría aplicar con los discursos racistas contra la población afroamericana, dada la experiencia histórica y aun el contexto actual, la regulación del discurso de odio en este respecto debería aplicarse de manera más estricta en EE.UU que en Alemania[57].

[57] Rosenfeld ilustra este punto analizando el tratamiento constitucional del *hate speech* en dos famosos casos de la jurisprudencia de la Corte Suprema de los Estados Unidos. Por un lado, el caso National Socialist Party of America v. Village of Skokie(1977) surgido a raíz de una marcha propuesta por los NeoNazis a través de Skokie, un suburbio en Chicago con una gran parte de su población judía. Frente a esto, las autoridades promulgaron legislación con el fin de evitar dicha marcha la cual los tribunales terminaron invalidando por violar la libertad de expresión de los NeoNazis. La autora explica que, debido a los factores contextuales de ese entonces en los Estados Unidos, la decisión del caso Skokie parece justificado, dado que tal como explica esta "la marcha real de los neonazis hizo mucho más para mostrar su aislamiento e impotencia que para promover su causa […] permitirles expresar su mensaje de odio probablemente contribuyó más a desacreditarlos que una prohibición judicial de su marcha [traducción propia]" (Rosenfeld 2003, 1538).

Sin embargo, la autora expresa que dicha jurisprudencia aplicada a otros contextos resulta bastante preocupante. Así, alude al caso R.A.V. v. City of St. Paul (1992) el cual se desató frente a la quema de una cruz llevada a cabo por extremistas blancos dentro del patio de una familia afroamericana. De esta manera, Rosenfeld argumenta que "debido a la naturaleza omnipresente del racismo y la larga historia de opresión y violencia contra los negros en los Estados Unidos, y dadas las aterradoras asociaciones evocadas por la quema de cruces, la situación en R.A. V. no puede equipararse con la

En definitiva, la propuesta de este apartado consiste en reflexionar acerca de la posibilidad de una regulación pública de las redes sociales en lo que respecta al discurso del odio que recepte un conjunto de principios normativos fundamentales a la luz de los estándares básicos de protección a nivel global de la libertad de expresión y a la vez los límites a la misma. Pero, imponiéndoles a las redes sociales el desafío de aplicar dichos estándares universales contemplando, a su vez, ciertas variables que hacen al caso en particular, como podrían ser: las diferencias culturales idiomáticas y prácticas lingüísticas de los diferentes lugares de los cuales puede provenir un mensaje de odio; la historia particular, las costumbres y el contexto del cual proviene el discurso y también de los destinatarios del mismo; el estatus de grupo (mayoritario, minoritario, reprimido, dominante, etc.)[58] del cual proviene el mensaje; y los objetivos e intereses que subyacen al discurso, entre otras.

4.3 El *censorship creep*

A la hora de pensar en una posible regulación pública de las redes sociales en cuanto a la difusión del *hate speech*, hay que tener especialmente en cuenta lo que Citron (2918) describe como "*censorship creep*". La autora utiliza esta expresión para referirse a una herramienta que es diseñada para lograr un propósito o resolver un problema en particular que gradualmente se extiende a otros usos o contextos. Aplicado al problema en cuestión, la autora aduce al riesgo que hay de la expansión de las políticas de expresión más allá de su objetivo original de limitar

involucrada en el caso Skokie […] aunque tanto la marcha propuesta en Skokie como la cruz en llamas en R.A. V. estaban destinados a incitar al odio sobre la base de religión y raza, respectivamente, sus efectos fueron bastante diferentes [traducción propia]" (Rosenfeld 2003, 1540). .

[58] En cuanto al grupo del cual proviene el mensaje de odio y al cual va dirigido, Rosenfeld argumenta que "es probable que el discurso racista de un miembro de una raza históricamente dominante contra los miembros de una raza oprimida tenga un impacto más severo que el discurso racista de los oprimidos racialmente contra sus opresores. Incluso si esto no justifica la regulación selectiva del discurso del odio, sí exige una mayor indulgencia cuando los oprimidos racialmente tienen la culpa, y que se tenga en cuenta como factor atenuante el hecho [...] que el discurso racista de un miembro de un grupo racial oprimido fue en respuesta al racismo perpetrado por miembros de la raza opresora [traducción propia]" (Rosenfeld 2003, 1566).

el discurso del odio. Así, menciona un comentario de Paul Bernal (2014) quien explica que "cuando se crea un sistema de censura para un propósito, se puede estar bastante seguro de que se utilizará para otros fines [traducción propia]" (Bernal 2014).

Dicho problema adquiere un tono aún más preocupante cuando se trata de las redes sociales que tienen un carácter transfronterizo. En este sentido, tal como explica la autora, mientras que el impacto en la censura que pueden llegar a tener las leyes nacionales está limitado por las fronteras geográficas, las limitaciones del *hate speech* en las redes sociales se da a través de los términos y condiciones de dichas plataformas los cuales, a menudo, se aplican a escala global. De esta manera, dado que los términos de servicio de los proveedores de redes sociales suelen ser los mismos en todo el mundo, las decisiones que las empresas privadas toman de eliminar o bloquear contenido por la violación de sus normas privadas aplicará en todos los países del mundo donde se tenga acceso a la plataforma en cuestión.

Así, explica Citron (2018) que los términos y condiciones de las empresas "podrían interpretarse para prohibir el discurso mucho más allá del discurso que incite al odio contra grupos vulnerables o del 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 critican a candidatos políticos. Podrían resultar en la suspensión global de los perfiles de Facebook de los activistas de derechos civiles [traducción propia]" (Citron 2018, 1050). Dicha propagación de la censura a escala global genera evidentes riesgos para el derecho fundamental de la libertad de expresión, en donde información esencial para el debate público sea eliminada universalmente a través de los términos de servicio.

Consecuentemente, la autora explica que para contrarrestar los riesgos del "*censorship creep*" las empresas deben, entre otras cosas, definir con claridad a que hacer referencia con el término ´discurso de odio´ y actuar con transparencia a la hora de limitar los contenidos por caer dentro de dicha definición. A continuación, abordaremos más en profundidad estos dos aspectos que se debería tener en cuenta en el diseño de una posible regulación.

a) <u>El concepto</u>

Cuando hacemos referencia a la limitación de la libertad de expresión frente al discurso de odio, un paso esencial radica en intentar discernir con la mayor precisión y claridad posible a qué hacemos referencia cuando hablamos de discurso de odio y qué es lo que no queda incluido dentro de dicho concepto. Esto ya que en última instancia cómo definamos al *hate speech* tendrá repercusiones directas en qué límites consideramos legítimos a un derecho tan fundamental como el que está en cuestión. Se requiere ser especialmente cuidadoso con la fina línea que hay entre el discurso de odio no amparado por la libertad de expresión y ciertos discursos de contenido crítico, disruptivo o "atrevido" que no solo son protegidos por dicha libertad sino que resulta fundamental que así lo sean.

Sin embargo, el principal problema al que nos enfrentamos es que no existe una definición universal de que se entiende cuando hablamos del discurso del odio, por el contrario, el concepto del discurso del odio resulta ser uno vago, maleable y subjetivo. Así, las legislaciones, jurisprudencia, doctrinarios y diferentes autores han propuesto diversos enfoques del concepto que varían según su contexto social, su cultura legal, el país de origen, sus antecedentes éticos y religiosos, entre otras cosas. Wenguang elude a esto sosteniendo que "a pesar de una plétora de estudios y análisis, el discurso del odio sigue siendo un fenómeno muy generalizado, difícil de definir, contextualizar y explorar con claridad. Las reglas sobre

contenido ilegal o dañino varían en diferentes países dependiendo de su historia social, ética, legal y religiosa [traducción propia]" (Wenguang 2018, 345).

Entre las diferentes definiciones que aportan los autores, Tulkens se refiere al *hate speech* como el "discurso que ataca intencionalmente a una persona o grupo por motivos de raza, etnia, género, discapacidad, orientación sexual, religión o cualquier otro criterio prohibido [traducción propia]" (Tulkens 2013). Por su parte, Marciani Burgos (2013) explica que el concepto de discurso de odio alude a "aquellas expresiones ofensivas dirigidas contra grupos humanos que han sido históricamente discriminados por motivos de género u opción sexual, raza, religión o situaciones similares. Lo que distingue al *hate speech* de otros conceptos semejantes (como las *fighting words*[59]), y lo vuelve problemático, es que las expresiones están dirigidas contra grupos y no contra individuos de forma particular, por lo cual no pueden subsumirse dentro de las figuras de la difamación, la calumnia o la injuria (Marciani Burgos 2013, 159). Por último, Mertsching define al *hate speech* como "la incitación pública a la violencia o al odio dirigida contra un grupo de personas o un miembro de dicho grupo definido por referencia a la raza, color, religión, ascendencia, origen nacional o étnico, sexo o género" (Mertsching 2018, 9).

Por otro lado, a nivel europeo nos encontramos con diferentes definiciones que arrojan luz sobre el significado del concepto. Un caso de esto es la Recomendación de 1997 del Comité del Consejo de Europa de Ministros sobre Discurso de Odio, en la cual se entiende como discurso del odio a "todas las formas de expresión que difundan, inciten, promuevan o

[59] Esta es una de las categorías desarrolladas por la jurisprudencia de la Corte Suprema de los Estados Unidos por la cual el discurso podría perder su protección constitucional bajo la Primera Enmienda. Explica Martín Herrera que "los *fighting words* son epítetos difamatorios susceptibles de despertar en el destinatario un sentimiento de emprendimiento de represalias en contra del emisor al ser emitidos directamente al oyente (…) [y] al no ser esencial para la expresión de una idea, y carecer de valor social, podrían llegar a ser prohibidas por violación a la paz social" (Martín Herrera 2021, 144). De todas maneras, el autor expresa que es importante tener en cuenta que, en más de seis décadas, la Corte no confirmó ni una sentencia bajo dicha categoría.

justifiquen el odio racial, la xenofobia, el antisemitismo u otras formas de odio basadas en la intolerancia, incluida la expresión intolerante mediante el nacionalismo agresivo y el etnocentrismo, discriminación y hostilidad contra minorías, migrantes y personas de origen inmigrante " (1997). Por su parte, a nivel judicial el Tribunal Europeo de Derechos Humanos ("TEDH") en el caso Müslum Günduz c. Turquía del 4 de diciembre de 2003, por ejemplo, definió al *hate speech* como "todas las formas de expresión que difunden, incitan, promueven o justifican el odio basado en la intolerancia, incluida la intolerancia religiosa"(TEDH 2003)[60].

A través de una lectura de las diferentes propuestas para definir el concepto de discurso del odio, en un primer momento nos podría llegar a parecer que todas aluden a lo mismo. Sin embargo, más allá de que todas las definiciones parezcan a primera vista muy parecidas, cuando las examinamos más en detalle podemos ver ciertas diferencias entre unas y otras. Así, en cuanto a la intencionalidad, algunos hacen referencia a discursos que "atacan" mientras que a otros les basta con que sea "ofensivo". Otros en vez aluden a la "incitación pública a la violencia o al odio" y otros a discursos que "difundan o promuevan el odio". En cuanto al contenido del discurso, la mayoría alude a motivos tales como raza, género, discapacidad, orientación sexual, religión, ascendencia, origen nacional o étnico. Si bien estos son los motivos que por lo general se mencionan, no todas las definiciones los incluyen en su totalidad. Por último, en referencia al destinatario del *hate speech*, algunos lo limitan tan solo a grupos y especifican que estos deben ser históricamente discriminados. Por su parte, otros explican que el discurso de odio también puede ser dirigido contra individuos sin necesidad de que estos pertenezcan a un determinado colectivo.

[60] Esta referencia es meramente ejemplificativa acerca de la postura de la TEDH, ya que la jurisprudencia del tribunal no ha ofrecido una respuesta concluyente acerca de qué tipo de mensajes constituyen el discurso de odio. Para un estudio profundizado al respecto ver: Diez Bueso (2020), "Discurso de odio en las redes sociales: la libertad de expresión en la encrucijada".

A raíz de este análisis podemos ver en definitiva lo dificultoso que puede llegar a resultar elaborar una definición jurídica del *hate speech* cuando de por sí ya nos basamos en un concepto tan vago y subjetivo como es el del odio. Asimismo, hay que ser especialmente precavidos con el hecho de que lo que puede llegar a ser sujeto de interpretación como discurso de odio bajo las definiciones previamente analizadas, puede ser a su vez justamente el objeto de lo que intenta proteger la libertad de expresión. Precisamente, la libertad de expresión surte sus mayores efectos a la hora de amparar al sátiro, al disidente, al hereje (Teruel Lozano 2017). Así, por ejemplo, un discurso sobre un asunto de interés público puede resultar "ofensivo" en base a ciertas imposiciones hegemónicas, pero justamente son estos discursos los que resulta fundamental proteger para tener la posibilidad de romper con ciertas imposiciones de grupos mayoritarias. En definitiva, hay ciertos discursos que pueden caer cerca de la definición de discurso de odio ya sea por consistir en ideas reprochables, ofensivas, desconcertantes o molestas, pero son curiosamente los especialmente protegidos bajo el alcance de la libertad de expresión[61].

Cuando pensamos en una posible regulación, la ambigüedad en la definición del *hate speech* puede resultar en un bloqueo excesivo que se extienda a contenido impopular o desfavorecedor, por ejemplo, al poder político de turno. Así Citron explica que "el *censorship creep* ocurre cuando las reglas del habla se basan en terminología ambigua. Sin pautas claras y ejemplos específicos, los términos vagos son vulnerables a revisión y expansión [traducción propia]" (Citron 2018, 1052). Esto podría incluso llevar al problema conocido como la "pendiente resbaladiza" según el cual dada la imposibilidad de trazar una línea clara entre que se considera discurso de odio y por ende ilegal y qué mensajes son legítimos y protegidos, una vez que la puerta a su regulación está abierta, gradualmente se podría permitir cada vez más

[61] *Ibid.,* 3

censura de todo tipo de discurso impopular, pero legítimo en fin, poniendo así en peligro a la libertad de expresión (Rosenfeld 2003). Es por esto que establecer una definición clara en la normativa que regule el contenido del discurso tiene como finalidad evitar el abuso por parte de las compañías privadas (ya sea movidas por razones privadas o a raíz de presiones externas, en específico estatales) silenciando discursos legítimos bajo la justificación de que en realidad se trata de discurso de odio, lo cual una definición ambigua les permitiría con facilidad.

Por otro lado, los conceptos jurídicos indeterminados conllevan el grave problema de la inseguridad jurídica, dentro de la cual, la falta de certeza legal en cuanto a las reglas que deben seguir las redes sociales podría incentivarlas a interpretar las reglas en su forma más estricta, aplicando de manera rigurosa las limitaciones del discurso, para asegurarse al máximo su exención de responsabilidad en caso de que el contenido sea considerado ilegal. Así, explican Sartor y De Azevedo Cunha que el derecho a la libertad de expresión de los usuarios de las redes sociales se puede ver obstaculizado ya que "es probable que se impida que cualquier información potencialmente controvertida llegue a ser accesible al público [traducción propia]" (Sartor y De Azevedo Cunha 2010, 376-377). En esta misma línea, explica Coche que en el marco de una consulta pública sobre la Directiva sobre el Comercio Electrónico (2000), "la mayoría de las partes interesadas (incluidos los intermediarios de Internet) opinaron que la eliminación excesiva de contenido se debe en parte a las incertidumbres legales que rodean el alcance y los términos de la exención de responsabilidad [traducción propia]" (Coche 2018, 7).

En suma, resulta necesario que las redes sociales en sus términos de servicio definan de manera clara y bien delimitada a qué se hace referencia con el término "discurso de odio", entre otras cosas, utilizando ejemplos específicos y explicando las razones de la restricción del discurso en estos casos (Isasi y Juanatey 2017). Según Citron, es necesario explicar las razones

de la prohibición de un tipo de contenido para que los usuarios tengan la información necesaria para comprender cuáles son sus derechos y sus responsabilidades en estas plataformas (Citron 2018). Una definición precisa del término "*hate speech*" es una herramienta clave para evitar que una regulación pública del discurso del odio a través de las redes sociales pueda derivar en censura. Por su parte, en la tarea de articular una definición lo suficientemente clara y delimitada del discurso de odio, las redes sociales podrían trabajar en conjunto con grupos que representen a múltiples sectores de la población, entre ellos, grupos defensores de libertades civiles, de derechos humanos, representantes políticos y académicos especializados en la temática, para ayudar a las empresas a llegar a una definición que recepte diferentes voces de la población y respete la libertad de expresión pero a la vez los demás derechos que se ponen en juego frente al *hate speech.*

b) La transparencia, la rendición de cuentas y la posibilidad de apelar las decisiones.

El otro aspecto que hay que tener en cuenta para evitar el *censorship creep* es lo que Citron (2018) denomina la "opacidad". Esta explica que en lo que respecta a la regulación del discurso llevado a cabo por entes privados, el contenido considerado de odio es eliminado por fuera de los procesos gubernamentales formales (ya sea judiciales o administrativos). Así, cuando se presenta una denuncia por contenidos publicados por usuarios que se consideran como discursos de odio, todo el proceso de decisión acerca de la legitimidad del discurso en cuestión es estudiado a través del sistema de denuncia de una compañía privada en vez de un proceso gubernamental que está obligado a cumplir con ciertos estándares procesales. Esto no da garantía alguna a los usuarios de transparencia en el proceso de decisión sobre la legitimidad del contenido en cuestión, lo cual manifiestamente pone en peligro el derecho a la libertad de

expresión de los usuarios, tanto de quien publicó el contenido como de los receptores del mismo[62].

De esta manera, un punto fundamental a incorporar en una posible regulación es establecer estándares de transparencia y rendición de cuentas por parte de las redes sociales a los usuarios en cuanto a la regulación del contenido en línea llevado a cabo por los intermediarios. En esta línea, Wenguang (2018) considera necesario que se lleve a cabo una apertura en donde las políticas de contenido de las compañías privadas y las medidas de monitoreo, filtrado y eliminación que lleven a cabo sean divulgadas y explicadas con especial claridad. Así, será de gran relevancia que las plataformas expliquen cuál es el contenido que está prohibido y cuales son los criterios que aplicarán para analizar los contenidos - y en su caso bloquearlos o eliminarlos-.

Asimismo, para contribuir en la transparencia Wenguang (2018) también propone establecer un régimen de rendición de cuentas tanto hacia el usuario que llevó a cabo la denuncia del contenido como para el usuario cuya publicación fue notificada o bloqueada, mediante el cual se les explique a los usuarios, en base a las políticas de la red social, el por qué de la decisión respecto al contenido en cuestión. En cuanto a la promoción de la transparencia y la rendición de cuentas, el Código de Conducta ha sido objeto de críticas. Así, explica Coche que "si bien el Código establece que promueve la transparencia, solo lo hace fomentando la publicación de informes de transparencia. En las dos últimas revisiones periódicas, no se prestó atención a la existencia de medidas de transparencia hacia los usuarios

[62] Sobre el derecho a la información como contracara del derecho a la libertad de expresión, la Corte Europea de Derechos Humanos, expresó en el caso Ligens v. Austria (1986) que "la prensa no solo tiene la tarea de difundir tales informaciones e ideas: el público también tiene derecho a recibirlas [...] La libertad de prensa ofrece al público uno de los mejores medios para descubrir y formarse una opinión sobre las ideas y actitudes de los líderes políticos. De manera más general, la libertad de debate político está en el centro mismo del concepto de sociedad democrática [traducción propia]"

finales cuya publicación había sido notificada y / o eliminada [traducción propia]" (Coche 2018, 8).

Por su parte, en particular es necesario establecer un sistema de transparencia riguroso para el caso de las solicitudes de eliminación de contenido relacionados con contenido o debate político[63]. En específico, si dichos pedidos son hechos por parte de los gobiernos, ya sea de manera directa o a través de organizaciones u otros actores que actúen en su representación. Así, Citron (2018) explica que las empresas privadas deberían proporcionar informes de transparencia especiales detallando los esfuerzos estatales por censurar los contenidos en línea, permitiendo de esta manera una conversación pública sobre la cesura en donde los usuarios puedan estar al tanto de los intentos gubernamentales de censurar contenidos mediante compañías privadas. A su vez, el autor argumenta que estas solicitudes llevadas a cabo por los gobiernos deben estar sujetas a una revisión rigurosa. El mismo sostiene que "cuando las autoridades gubernamentales buscan suprimir el discurso bajo los términos de servicio, los moderadores de contenido deben ver las solicitudes con una presunción en contra de la eliminación, o al menos con una buena dosis de escepticismo. (...) [estos] deben recibir capacitación sobre el *censorship creep*, incluidos los esfuerzos gubernamentales pasados y presentes para silenciar a los críticos. La formación debe centrarse en cómo distinguir los discursos de incitación al odio o el material terrorista prohibido del contenido de interés periodístico" (Citron 2018, 1066). Por último, cualquier decisión relacionada con solicitudes gubernamentales de esta índole deberían estar detalladas en informes especiales en donde

[63] Sobre este punto, la Corte Europea de Derechos Humanos, en el caso Incal v. Turkey (1998), sostuvo que al evaluar el margen para restringir el discurso político o sobre asuntos de interés público "los límites de la crítica permisible son más amplios con respecto al gobierno que con respecto a un ciudadano privado, o incluso a un político. En un sistema democrático, las acciones u omisiones del gobierno deben estar sujetas al escrutinio no solo de las autoridades legislativas y judiciales, sino también de la opinión pública [traducción propia]".

quienes tomen las decisiones al respecto deban explicar de manera aún más detallada cuales fueron los criterios tenidos en cuenta.

Por último, para contribuir a la transparencia de las decisiones llevadas a cabo por la red social y en aras de proteger el derecho a la libertad de expresión de los usuarios, otra de las medidas que podría establecer la regulación es que las plataformas implementen un proceso mediante el cual se habilite a los usuarios cuyo contenido fue bloqueado por infringir las políticas relacionadas al *hate speech*, a recurrir dicha decisión. En este sentido, Gascón Marcen propone que "se dé información a las personas que notifican discurso de odio sobre las medidas tomadas y que se creen mecanismos sencillos para las personas que consideren que sus contenidos no son ilícitos para que puedan recurrir la decisión" (Gascón Marcen 2019, 83).

Un ejemplo paradigmático de esto es el "Consejo Asesor de Contenido" (en adelante, el "Consejo"), creado a los fines de ayudar a Facebook a abordar las problemáticas surgidas frente a la libertad de expresión y sus políticas de contenido de odio. Tal como explica el Consejo en su página oficial[64], en orden a cumplir con su finalidad de proteger la libertad de expresión, el mismo emplea la independencia de su criterio para tomar decisiones autónomas basadas en los principios de contenido de Facebook. En este sentido, la labor del Consejo consiste en someter a revisión las decisiones de contenido tomadas por la red social (por ahora elige un número limitado de casos emblemáticos[65]) y decidir, de manera vinculante para la

[64] A través del siguiente link se puede ingresar a la página oficial del "Consejo Asesor de Contenido" para más información del mismo: https://oversightboard.com/

[65] Uno de estos casos emblemáticos fue la revisión por parte del Consejo acerca de la decisión tomada por Facebook en enero de 2021 de restringir el acceso de manera indefinida del entonces presidente de los Estados Unidos Donald Trump a sus cuentas de Facebook e Instagram, en el marco de ciertas publicaciones llevadas a cabo por el líder el día que se llevó a cabo la toma del Capitolio en Washington.

Si bien la Junta ratificó la decisión de la red social considerando que dichas publicaciones eran violatorios de las Normas Comunitarias de Facebook que prohíben apoyar o elogiar los eventos que estaban ocurriendo en el Capitolio, el Consejo resolvió que Facebook debía revisar su decisión en

última, si rectificarlas o revertirlas de acuerdo con las políticas y los valores de Facebook. Dicha decisión será vinculante salvo que cumplir con la misma suponga infringir la ley. Por su parte, el consejo cuenta con miembros de distintas partes del mundo, conformando así un equipo con conocimientos y disciplinas diversas. Además, una vez tomada la decisión por parte del Consejo, se notifica a la persona que presentó la apelación y los fundamentos de la misma serán puestos a disposición de la apelante y también de la totalidad del público.

De esta manera, mediante los mecanismos de transparencia, de notificación, de rendición de cuentas y de apelación de las decisiones, se podría construir el camino hacia una regulación del discurso de odio *online* en donde los riesgos de censura se vean atenuados ampliamente.

4.4 ***Multi Stakeholder approach*** **y la censura como última** ***ratio***

Por su parte, considero que una regulación pública que establezca la responsabilidad a las plataformas de redes sociales en la lucha contra la multiplicación del discurso de odio en internet debería partir de un enfoque de múltiples partes interesadas (*multi stakeholders approach*). Explica Wenguang que "un enfoque de múltiples partes interesadas y una perspectiva de gobernanza podrían brindar posibilidades para abordar esos problemas complejos relacionados con las responsabilidades de los intermediarios por contenido ilegal de terceros [traducción propia]" (Wenguang 2018, 356).

cuanto a la proporcionalidad de la sanción. En este sentido, el Consejo sostuvo que no resultaba apropiado que Facebook impusiera una restricción indefinida en las cuentas en cuestión.

Para acceso completo al caso, ingresar a: https://oversightboard.com/decision/FB-691QAMHJ/

Así, la regulación debería establecer el deber de que en el diseño de sus políticas restrictivas del *hate speech* en los términos de servicio, las compañías privadas tengan en cuenta y logren un equilibrio entre los múltiples intereses que surgen de diferentes actores de la sociedad. Para esto, resulta importante la participación y colaboración de todas las partes interesadas, como lo podrían ser los intermediarios, el gobierno, los usuarios, organizaciones de la sociedad civil que representen a la comunidad en su conjunto y organizaciones no gubernamentales que luchan por causas tales como el racismo, la xenofobia, la intolerancia religiosa, etc.

Por otro lado, considero que también es de suma importancia que la regulación estatal incentive a las compañías privadas a que en el diseño de sus políticas de *hate speech* incluya otros mecanismos y alternativas de control del discurso de odio y que el bloqueo o eliminación de contenido sea utilizado como herramienta de última *ratio*. Así, sostiene Marciani Burgos que "tomando en cuenta los peligros derivados de la censura legal, esta debería ser una opción residual frente a otras alternativas que resulten menos invasivas por parte del Estado y que, más bien, reconozcan la agencia de los afectados y su capacidad para provocar respuestas y cambios" (Marciani Burgos 2013, 196). En más, no solo por los riesgos de censura, sino que también muchas veces, el bloqueo de contenidos *online* puede parecer la manera más eficaz de luchar contra el discurso del odio, pero sin embargo puede terminar derivando en otros problemas como el de socavar investigaciones policiales en curso[66] y también terminar convirtiendo en mártires a aquellos sujetos cuyos mensajes han sido eliminados por la atención mediática que despiertan este tipo de medidas restrictivas del discruso (Gascón Marcen 2019).

[66] Explica Citron que "hay otros costos más allá del ámbito de la libre expresión. La eliminación del discurso extremista puede dificultar que las fuerzas del orden hagan su trabajo. Las investigaciones de terrorismo a menudo se basan en pistas que se dejan en la actividad de las redes sociales. Por lo tanto, puede ser difícil investigar el terrorismo potencial si las pruebas en línea se eliminan de inmediato [traducción propia]" (Citron 2018, 1061).

Asimismo, como mencionamos anteriormente la libertad de expresión protege especialmente al sátiro, disidente o hereje, y por ende el Estado debería intervenir para promover activamente la mayor cantidad de discursos alternativos que puedan enfrentarse a los valores hegemónicos de la sociedad. Así, explica Teruel Lozano: "No creo que el Estado deba ser neutral, porque, en una democracia social (...) el Estado no es nihilista y debe defender los valores constitucionales pero, en materia de expresión, deberá hacerlo no desde la censura, sino a través de otras políticas" (Teruel Lozano 2017, 189).

Al diseñar los términos de servicio las compañías deben tomar conciencia de que el fenómeno del odio ocurre en un contexto tanto social como histórico, y que por lo tanto, censurar o prohibir los mensajes de odio por sí solo no hará desaparecer dicho fenómenos ni evitará las consecuencias que este trae (Díaz 2017). De esta manera, es importante que la regulación del *hate speech* en las redes sociales ofrezcan diversas alternativas al control del último que serán utilizadas de manera primaria. Estas deberían hacer foco, entre otros, en la educación de los usuarios de las redes sociales para fomentar el respeto a la diversidad y los derechos humanos, al igual que educar en el valor de la tolerancia frente a las diferentes culturas, religiones o creencias (Carrillo Donaire 2015)[67]. Asimismo, se deberían también

[67] Este enfoque que se centra en educar a los usuarios de las redes en vez de eliminar la información que consumen fue especialmente tomado por los intermediarios de Internet a partir de la masiva desinformación que se generó como consecuencia de la pandemia del Covid-19. En tal sentido, explica la Organización de la Naciones Unidas para la Educación, la Ciencia y la Cultura ("UNESCO") que a medida que el Coronavirus se expandió por el mundo, de igual manera lo hizo la información falsa la cual dificultó a la gente encontrar fuentes fidedignas que los informaran acerca del virus.

En tal sentido, el informe explica que las redes sociales se han convertido en grandes focos de desinformación. Frente a esto, las compañías mediante una declaración conjunta anunciaron su compromiso por combatir este problema, en específico, orientando a los usuarios hacia información oficial acerca del virus. En dicho comunicado, las plataformas anunciaron: "estamos trabajando en estrecha colaboración en los esfuerzos de respuesta al COVID-19 [...] combatimos conjuntamente el fraude y la información errónea sobre el virus, aumentamos el contenido autorizado en nuestras plataformas y compartimos actualizaciones críticas en coordinación con las agencias gubernamentales de atención médica de todo el mundo. Invitamos a otras empresas a unirse a nosotros mientras trabajamos para mantener nuestras comunidades saludables y seguras [traducción propia]" (@Googlepubpolicy, 16 de marzo de 2020).

centrar en el en el *counter speech*[68] al discurso del odio mediante la creación de mecanismos que promuevan la contra narrativa, ya sea para promover la no discriminación y la tolerancia, entre otros, como también la réplica de los grupos vulnerables víctimas del odio y, en última instancia, la reconciliación.

En lo que respecta a priorizar otras alternativas antes que la censura de contenidos, debemos también tener en cuenta que el derecho a la libertad de expresión no solo portege a la persona creadora del contenido que está en jaque, sino que en la otra cara de la moneda se encuentra el derecho de las personas a poder recibir dicha información[69]. Las redes sociales no solo expandieron enormemente la capacidad de las personas para participar en el discurso, sino que también han mejorado la capacidad del público para escuchar y comprender (Ardia 2010). Por ende, en lo que concierne al *hate speech*, resulta necesario que la gente pueda tener acceso y conocimiento a las ideas u opiniones de odio, para poder cuestionar y en su caso repudiar y

Para más información acerca de la tensión entre el derecho de libertad de expresión y las *fake news*, escuchar el podcast de Federico Carestia, titulado "Fake News y Libertado de Expresión" disponible en: https://open.spotify.com/episode/07gxktk7kEbOXebImOYOqJ?si=OVD7Q7OvSmSkcjdzAhcadA&dl_branch=1

[68] En este sentido, en el marco de la iniciativa para combatir el discurso de odio y extremista en línea llevada a cabo por Facebook, Sheryl Sandberg - Directora Operativa de Facebook- sostuvo que "la mejor cura para las malas ideas son las buenas ideas [...] El mejor remedio para el odio es la tolerancia. El *counter speech* es increíblemente fuerte y requiere tiempo, energía y coraje" (Sandberg 2016). Mediante dicha iniciativa llamada *"Online Civil Courage Initiative",* se animó a las personas a compartir sus historias e ideas mediante el hashtag #civilcourage y así mostrar su apoyo a la iniciativa.

Otro ejemplo muy interesante sobre el uso del *counter speech* es el proyecto llevado a cabo por Jigsaw, una unidad de expertos dentro de Google que crea tecnología para solucionar las amenazas a las sociedades abiertas. En este sentido, Jigsaw ha desarrollado un programa que a través de la combinación entre algoritmos de publicidad y la plataforma Youtube apunta a disuadir a unirse a ISIS a los aspirantes a reclutas. En su página oficial, Jigsaw explica que "el método de redireccionamiento utiliza herramientas de orientación de AdWords y videos seleccionados de YouTube cargados por personas de todo el mundo para enfrentar la radicalización en línea. Se centra en el segmento de la audiencia de ISIS que es más susceptible a sus mensajes y los redirige hacia videos seleccionados de YouTube que desacreditan los temas de reclutamiento de ISIS". Para más información, ingresar al siguiente link: https://redirectmethod.org/

[69] *Ibid.,* 63

contrarrestar dichas opiniones (Citron 2018).[70] Como lo plantea Aryeh Neier, "la libertad de expresión en sí misma sirve como el mejor antídoto contra las doctrinas venenosas de quienes intentan promover el odio [traducción propia]" (Citron 2018, 1058).

4.5 Protección de la libertad de expresión a la altura de las cartas fundamentales

Por último, al colocar a las compañías privadas de redes sociales a cargo de la aplicación de la ley en lo que respecta al discurso de odio en línea, emerge el riesgo de que el derecho a la libertad de expresión se termine rigiendo según los estándares que establecen los términos de servicio de la empresa, los cuales podrían no llegar a garantizar el mismo grado de protección que garantizan a dicho derecho fundamental los diferentes instrumentos de derechos humanos (Coche 2018). En este sentido, bajo una regulación estatal que responsabilice a las redes sociales por el contenido publicado por terceros, se podría terminar eliminando contenido por incumplir los términos de servicio de la compañía pero que constituía discurso legítimo según los estándares legales que protegen el derecho fundamental de la libertad de expresión. Dicha preocupación se dio en torno al Código de Conducta, según el cual la remoción de contenido se basa principalmente en los términos de servicio y solo de manera secundaria en la legislación nacional.[71] En este sentido, el Consejo de Europa expresó como una amenaza al Estado de derecho[72] el hecho de que los términos y condiciones diseñados por las redes sociales

[70] Como expresa Barata, en la entrevista realizada para la revista Xataca y en relación al bloqueo de la cuenta de Trump en Twitter, "En algunos casos hay que dar especial visibilidad a lo que dicen los políticos, para saber sus propios juicios. Si se silencian, será más difícil poder juzgarlos. Si un político dice un disparate, es bueno que sepamos lo que dice" (Barata 2021).

[71] Tal como fue mencionado previamente, a través del Código de Conducta las redes sociales se comprometen a evaluar las solicitudes de eliminación en función de sus normas internas, y solo de manera supletoria de ser necesario, de las leyes nacionales que transpongan la Decisión marco 2008/913 / JAI.

[72] El Consejo de Europa describió al estado de derecho como "un principio de gobernanza por el cual todas las personas, instituciones y entidades, públicas y privadas, incluido el estado mismo, son responsables de las leyes que se promulgan públicamente, se aplican por igual, se adjudican y se

no estén de acuerdo con las normas internacionales de derechos humanos. Coche (2018) ilustra este punto explicando lo que sucedió con el Código de Conducta. El mismo plantea que "este instrumento coloca las responsabilidades de ejecución en manos de empresas privadas (...) Los peligros derivados de dicha práctica pueden ilustrarse en el último informe semestral de Twitter (2017), en el que indica que desde julio de 2017 hasta diciembre de 2017, se suspendieron 274,460 cuentas debido a actividades relacionadas con el terrorismo en violación de los términos y servicios de la empresa […] el 19% resultó en la eliminación de contenido debido a violaciones de los términos de servicio (TOS) y el 10% en contenido retenido en un país en particular de acuerdo con las leyes locales" (Coche 2018, 2). Así, una de las mayores preocupaciones acerca del Código de Conducta como explica Gascón Marcen (2019), es que parecería que el Código rebaja a un segundo plano a la ley, tras el papel principal que juegan las compañías privadas proveedoras de las redes sociales quienes, a partir de su implementación arbitrarias de sus términos de servicio generan grandes "riesgos para la libertad de expresión ya que el contenido legal pero controvertido puede eliminarse como resultado de este mecanismo de eliminación voluntario e irresponsable" (Gascón Marcen 2019, 71).

Es por esto que, uno de las principales cuestiones a tener en cuenta a la hora de reflexionar en una regulación estatal que responsabilice a las redes sociales como intermediarias por el contenido publicado por terceros es el hecho de que los términos de servicio sean demasiado amplios, amenazando el estado de derecho actual y los niveles de protección legales que se le otorga a la libertad de expresión. En vistas de esto, se debe tener especial cuidado en que dicha regulación pública no incentiva a las redes sociales al bloqueo o eliminación de contenidos lícitos. Para esto, se debería reflexionar en la posibilidad de una regulación estatal que establezca ciertos lineamientos y estándares mínimos a la hora del diseño

adjudican de manera independiente. coherente con las normas y estándares internacionales de derechos humanos [traducción propia]" (Consejo de Europa 2014, 10).

de las políticas relacionadas al *hate speech* de las redes sociales para la protección de la libertad de expresión de acuerdo con las normas internacionales de derechos humanos. Así, resulta relevante que la regulación pública del discurso del odio en las redes sociales debería hacer especial énfasis en que las políticas de contenido de odio protejan al derecho fundamental de la libertad de expresión como lo está protegido a nivel legal a través de las cartas de derechos fundamentales.

5. Conclusión

Las redes sociales cumplen con un rol fundamental al servicio de la libertad de expresión, constituyéndose en plataformas para la comunicación libre de censura estatal. Sin embargo, el amplio alcance de las redes sociales generado por su potencial de difusión, sumado a la dificultad de monitorear y eliminar las publicaciones que se difunden y el anonimato que genera una sensación de impunidad a los usuarios, las convierte en instrumentos ideales para la propagación del *hate speech.* Esto torna especialmente relevante al tomar conciencia de las consecuencias a las que puede derivar la proliferación *online* del discurso de odio, las cuales pueden ser tanto en línea como fuera de línea y pueden ir desde el daño emocional y psicológico de los individuos objeto de amenazas y acoso, hasta la perpetuación de estereotipos discriminatorios que contribuyen a la marginalización de ciertos colectivos y la segregación de la sociedad.

Es por esto que, en vistas a la dimensión que ha adquirido el discurso de odio a través de las redes sociales y a los hechos de violencia fuera de línea que se pueden desencadenar a partir de esto, el presente trabajo se propuso reflexionar acerca de la necesidad de algún tipo de regulación para controlar este fenómeno. Como se advirtió, la regulación del discurso de odio plantea especiales complejidades que nos hacen cuestionar y repensar los límites de la libertad de expresión. Dado que son justamente los mensajes disidentes, molestos u ofensivos

los que resulta necesario y deseable difundir y hacer visible dentro del debate público, resulta fundamental mirar con recelo cualquier tipo de restricción a expresiones catalogadas como "de odio". Sin embargo, el derecho a la libertad de expresión no es absoluto, sino que encuentra sus límites frente otros bienes jurídicos y valores relevantes para las democracias occidentales.

Por ende, nos encontramos frente a la delicada y controversial tarea de balancear y sopesar entre la protección de este pilar del sistema democrático y las expresiones de odio que, frente a la escala masiva que pueden adquirir a través de las redes sociales, podrían atentar contra la dignidad humana y la convivencia social. A la hora de pensar en una posible regulación que limite el *hate speech* sin poner en riesgo el derecho fundamental que está en juego, se planteó pensar en tres diferentes esquemas de regulación, de las cuales se abordaron específicamente dos. Por un lado, se analizó la regulación pública del contenido publicado en las redes sociales, para lo cual se abordaron los dos principales paradigmas de regulación. Por el otro lado, se abordó la posibilidad de que sean las propias compañías las que establezcan, mediante sus políticas de contenido de odio, qué tipo de discursos permiten difundir en sus plataformas.

En este sentido, se explicó cómo si bien cuando hacemos referencia a la regulación privada del *hate speech* en principio no cabe hacer referencia al derecho de la libertad de expresión - ya que esta protege a los individuos frente al poder de censura por parte del Estado, y no de los privados- el rol de las redes sociales como reguladores de la expresión *online* torna relevante al reconocer que estas ya no diseñan sus políticas de manera independiente y movidas por las necesidades del mercado, sino que - por el contrario- hoy en día son el resultado de presión por parte de los poderes públicos. Es por esto que torna difícil - y un tanto ingenuo- pensar en la posibilidad de que, en caso de sostener que la regulación estatal del contenido *online* resulte peligrosa para la libertad de expresión, la otra opción sea dejar que la cuestión

se regule por las fuerzas del mercado y en base a las decisiones libres e independientes de las compañías privadas. Por el contrario, dado que los cambios de política de contenido llevados a cabo por las redes sociales son el resultado de presiones estatales, que la regulación del *hate speech* se esconda tras el ropaje de lo privado pone aún más en riesgo el derecho fundamental en cuestión.

En definitiva, este trabajo es una contribución que intenta evidenciar la actual e inevitable necesidad de reflexionar sobre una posible alternativa de regulación estatal que aborde la problemática de la proliferación del *hate speech* en las redes sociales. Pero esto, desde una perspectiva que tenga como eje central la protección de la libertad de expresión, máxime, al considerar que dichas plataformas conforman el espacio ideal para la expresión de todo tipo de ideas sin control estatal y que, tras el ropaje del discurso de odio, los Estados podrían abusar de dicha regulación y justificar con ella la censura estatal. Por ello, como puntapié inicial se aspiró proponer ciertos estándares básicos – que para nada pretenden ser exhaustivos de la cuestión- que se deberían tener en cuenta a la hora de diseñar una regulación estatal que imponga a las redes sociales la obligación de abordar el discurso de odio difundido a través de sus plataformas, pero a su vez protegiendo el derecho fundamental en juego.

Más allá de los puntos propuestas en aras de pensar el diseño de una regulación que proteja el derecho a la libertad de expresión, es menester para futuros trabajos profundizar en cómo se podría llevar a cabo la aplicación práctica de una regulación de estas características. En especial, dado el carácter transnacional de la cuestión resulta relevante reflexionar sobre la efectiva aplicabilidad de la regulación y sobre un posible sistema de sanciones a las redes sociales que contribuya a reforzar la vinculatoriedad de dicha norma. Por su parte, un tema de gran importancia a analizar es la manera en que las redes sociales podrían llevar un efectivo y eficiente control sobre el contenido difundido mediante las plataformas para poder detectar el

discurso de odio, en específico, teniendo en cuenta las limitaciones que aún presenta la inteligencia artificial a estos fines[73].

[73] Al respecto, Cabo Isasi y Juanatey explican que "Facebook ha desarrollado un potente sistema de inteligencia artificial, denominado DeepText, supuestamente capaz de analizar varios miles de posts por segundo en más de 20 lenguas. Sin embargo, a día de hoy no parece que su funcionamiento esté siendo un éxito. Por muy sofisticados que sean los algoritmos en los que se basan este y otros filtros informáticos desarrollados por otras redes sociales, tienden a cometer errores de bulto y no suelen ser capaces, por ejemplo, de discernir cuando se está haciendo un uso crítico o de denuncia de una determinada expresión insultante" (Cabo Isasi y Juanatey 2017, 22).

6. Referencias

- *Abrams v. United States,* 250 U.S. 616 (1919).
- Alex Cabo Isasi y Ana García Juanatey. 2017. "El discurso del odio en las redes sociales: Un estado de la cuestión". Barcelona: Ajuntament de Barcelona Progress Report.
- Alkiviadou, Natalie.2018. "The Legal Regulation of Hate Speech: The International and European Frameworks". *Croatian Political Science Review,* no.4 (55): 203-229.
- Almeida, Carlos S. 2020. "Eliminar mensajes de odio sin pasar por un juez es un plan peligroso: las dudas que plantea la proposición de Unidas Podemos". Entrevista por Enrique Pérez. *Xataka,* 5 Noviembre 2020. https://www.xataka.com/legislacion-y-derechos/eliminar-mensajes-odio-pasar-juez-propuesta-peligrosa-dudas-que-plantea-proyecto-no-ley-unidas-podemos
- Badawy, Adam, y Ferrara, Emilio. 2018. "The Rise of Jihadist Propaganda on Social Networks". *Journal of Computational Social Science,* 1(2), 453-470.
- Balkin, Jack M. 2017. "Free speech in the algorithmic society: Big data, private governance, and new school speech regulation". *UCDL Rev., 51,* 1149-1210.
- Barata, Joan. 2021. "Trump, Twitter y el gran debate sobre la "censura": quién tiene el poder para marcar qué puede decirse en las redes sociales." Xataca, 18 Enero de 2021. https://www.xataka.com/legislacion-y-derechos/trump-twitter-gran-debate-censura-quien-tiene-poder-para-marcar-que-puede-decirse-redes-sociales
- Bernal, Paul. 2014. "Censorship and surveillance… " *Paul Bernal's Blog Privacy, Human Rights, Law, The Internet, Politics and more* (blog). 25 de septiembre de 2014. https://paulbernal.wordpress.com/2014/09/25/censorship-and-surveillance/
- Boix Palop, Andrés. 2016. "La Construcción de los límites a la libertad de expresión en las redes sociales". *Revista de Estudios Políticos,* no. 173: 55-112.

- Carrasco, Sergio. 2020. “La UE quiere multar a las redes sociales por los contenidos que considere ilegales: la libertad de expresión puede estar en peligro”. Entrevista por Antonio Sabán. *Genbeta,* 3 Noviembre 2020. https://www.genbeta.com/redes-sociales-y-comunidades/ue-quiere-multar-a-redes-sociales-contenidos-que-consideren-ilegales-libertad-expresion-puede-estar-peligro
- Carrillo Donaire, Juan Antonio. 2015. “Libertad de Expresión y discurso del odio religioso: la construcción de la tolerancia en la era postsecular”. *Revista de Fomento Social,* no. 70: 205-243.
- Citron, Danielle Keats. 2018. “Extremist Speech, Compelled Conformity, and Censorship Creep”. *Notre Dame Law Review*, no. 3 (93): 1035-1072.
- Clark, Liat. 2016. “Facebook and Twitter must tackle hate speech or face new laws”. *Wired UK,* 5 de diciembre de 2016. https://www.wired.co.uk/article/us-tech-giants-must-tackle-hate-speech-or-face-legal-action
- Coche, Eugénie. 2018. “Privatised Enforcement and the Right to Freedom of Expression in a World Confronted with Terrorismo Propaganda Online ”. *Internet Policy Review,* no. 7.4.
- Comisión Europea. 2016. "European Commission And IT Companies Announce Code Of Conduct On Illegal Online Hate Speech". 31 de mayo de 2016. Accedido el 26 de junio de 2021. https://ec.europa.eu/commission/presscorner/detail/en/IP_16_1937.
- Comité del Consejo de Europa de Ministros sobre Discurso de Odio. 1997. Recomendación de 1997. (Citado en: Alkiviadou, Natalie.2018. “The Legal Regulation of Hate Speech: The International and European Frameworks”. *Croatian Political Science Review*, no.4 (55): 203-229.
- Communications Decency Act. 47 U.S.C § 230. (1996).

- Independent International Fact-Finding Mission on Myanmar. 2018, A/HRC/39/CRP.2 "Report of the detailed findings of the Independent International Fact-Finding Mission on Myanmar". Disponible en: https://undocs.org/A/HRC/39/CRP.2
- Decisión Marco 2008/919/JAI.
- Copyright Directive, Article 17th § (2019).
- *Delfi AS v. Estonia.* 2015, App no. 64569/09.
- Díaz, Marianne. 2017. "El odio y los límites a la libertad de expresión". *Derechos Digitales. Derechos Humanos y Tecnología en América Latina*, 8 de septiembre, 2017. https://www.derechosdigitales.org/11421/el-odio-y-los-limites-a-la-libertad-de-expresion/
- Electronic Frontier Foundation. s.f. "Section 230 of the Communications Decency Act". Accedido el 26 de junio de 2021. https://www.eff.org/issues/cda230
- Fiss, Owen. 1996. "The Irony of Free Speech". *Harvard University Press.*
- *Fressoz and Roire v. France*. 1999, App no. 29183/95.
- Gascón Marcen, Ana. 2019. "La lucha contra el discurso de odio en línea en la Unión Europea y los intermediarios de internet". En *Libertad de expresión y discurso de odio por motivos religiosos.* Licregdi, 64-86.
- Goyret, Lucas. 2021. "La dictadura de Nicolás Maduro arrecia la censura en Venezuela: ahora va por el control de las redes sociales." *Infobae*, 23 de abril de 2021. https://www.infobae.com/america/venezuela/2021/04/23/la-dictadura-de-nicolas-maduro-arrecia-la-censura-en-venezuela-ahora-va-por-el-control-de-las-redes-sociales/
- *Hertel v. Switzerland*. 1998, App no. 25181/94.
- *Incal v. Turkey*. 1998, App no 22678/93.
- Kimani, Mary. 2015. "RTLM: the Medium that Became a Tool for Mass Murder". *The Media and the Rwanda Genocide*, 110. Pluto Press.

- Korff, Douwe. 2014. “The rule of law on the internet and in the wider digital world”. *Issue Paper Published by the Council of Europe Commissioner for Human Rights, Council of Europe*.
- *Ligens v. Austria*. 1986, App no. 9815/82.
- Marciani Burgos, Betzabé. 2013. “El lenguaje sexista y el hate speech: un pretexto para discutir sobre los límites de la libertad de expresión y de la tolerancia liberal”. *Revista Derecho del Estado,* no. 30: 157-198.
- Martín Herrera, David. 2021. *Extreme speech y libertad de expresión análisis de la jurisprudencia constitucional de la Corte Suprema estadounidense*. Madrid: Dykinson, S.L., 2018.
- *Matal v. Tam*. 2017, 582 U. S.
- Mertsching, Saskia. 2018. “Online hate speech regulation in the European Union and Germany”. Tesis de Maestría, Universidad de Derecho de Oslo.
- Pérez, Enrique. 2020. “Eliminar mensajes de odio sin pasar por un juez es un plan peligroso: las dudas que plantea la proposición de Unidas Podemos”. *Xataka*, 5 Noviembre 2020.https://www.xataka.com/legislacion-y-derechos/eliminar-mensajes-odio-pasar-juez-propuesta-peligrosa-dudas-que-plantea-proyecto-no-ley-unidas-podemos
- Pérez, Enrique. 2021. “Trump, Twitter y el gran debate sobre la "censura": quién tiene el poder para marcar qué puede decirse en las redes sociales”. *Xataka*, 18 Enero 2021. https://www.xataka.com/legislacion-y-derechos/trump-twitter-gran-debate-censura-quien-tiene-poder-para-marcar-que-puede-decirse-redes-sociales
- *Republic of Poland v European Parliament and Council of the European Union.* 2019. Case C-401/19.
- Rivera, Julio César (h.). 2006. “El derecho a la intimidad como límite a la publicación de sentencias judiciales. Análisis crítico de la doctrina sentada en los casos "P.A. c/Arte

Gráfico Editorial Argentino SA" y "P.A. c/Diario La Prensa" desde una perspectiva constitucional". *Revista de Derecho Privado y Comunitario*.

- *Rodríguez, María Belén c/ Google Inc. y otro s/ daños y perjuicios*. 2014. R.522.XLIX
- Rosenfeld, Michel. 2003. "Hate Speech in Constitutional Jurisprudence: A Comparative Analysis". *Cardozo Law Review*, no. 24: 1523-1568.
- Sabán, Antonio. 2020. "La UE quiere multar a las redes sociales por los contenidos que considere ilegales: la libertad de expresión puede estar en peligro". *Genbeta*, 3 Noviembre 2020.https://www.genbeta.com/redes-sociales-y-comunidades/ue-quiere-multar-a-redes-sociales-contenidos-que-consideren-ilegales-libertad-expresion-puede-estar-peligro
- Sartor, Giovanni, y De Azevedo Cunha, Mario Viola. 2010. "The Italian Google-case: Privacy, freedom of speech and responsibility of providers for user-generated contents." *International Journal of Law and Information Technology,* no. 18 (4): 356-378.
- Sevanian, Andrew, M. 2014. "Section 230 of the communications decency act: A good samaritan law without the requirement of acting as a good samaritan". *UCLA Entertainment Law Review*, no. 21 (1): 121-145.
- Smith, Dave. 2016. "Facebook has launched a new campaign against hate speech". *Insider.* 19 de enero de 2016. https://www.businessinsider.com/facebook-online-civil-courage-initiative-2016-1
- *Sürek v. Turkey (No 1).* 1999, App no. 26682/95.
- Teruel Lozano, Germán. 2017. "Expresiones intolerantes, delitos de odio y libertad de expresión: un difícil equilibrio". *Revista Jurídica*, no. 36: 185-196.
- Teruel Lozano, Germán. 2020. "Eliminar mensajes de odio sin pasar por un juez es un plan peligroso: las dudas que plantea la proposición de Unidas Podemos". *Xataka.* 5 de noviembre de 2020. https://www.xataka.com/legislacion-y-derechos/eliminar-mensajes-odio-pasar-juez-propuesta-peligrosa-dudas-que-plantea-proyecto-no-ley-unidas-podemos

- *The prosecutor v. Ferdinand Nahimana, Jean-Bosco Barayagwiza, Hassan Ngeze*. 2003, Case No. 1CTR-99-52-T.
- Tórtora Aravena, Hugo. 2010. “Las limitaciones a los derechos fundamentales”. *Estudios constitucionales,* no. 2: 167-200.
- Tulkens, Françoise. 2013. “The Hate Factor in Political Speech. Where Do Responsibilities Lie?”. En *Report of the Council of Europe Conference*.
- UNESCO. 2020. "Periodismo, Libertad De Prensa Y COVID-19". https://en.unesco.org/sites/default/files/unesco_covid_brief_es.pdf.
- Uttamchandani, Rahul . 2020. “La libertad de expresión en Internet: La sección 230 de la Communications Decency Act”. Última modificación 29 de octubre de 2020. https://www.legalarmy.net/la-libertad-de-expresion-en-internet-la-seccion-230-de-la-communications-decency-act/
- Vega, Guillermo. 2020. “Google, Facebook y Twitter defienden ante el Senado de EE UU su papel en Internet.” *El País*, 28 de octubre de 2020. https://elpais.com/tecnologia/2020-10-28/google-facebook-y-twitter-defienden-ante-el-senado-de-ee-uu-su-papel-en-internet.html
- Vuarambon, Nicole. S.f. “El paradigma de las redes sociales: Entre la libertad de expresión y la censura”. *DemoAmLat.* Accedido el 19/6/2021. *https://www.demoamlat.com/el-paradigma-de-las-redes-sociales-entre-la-libertad-de-expresion-y-la-censura/*
- Wenguang, Yu. 2018. “Internet Intermediaries´ Liability for Online Illegal Hate Speech”. *Frontiers of Law in China,* no. 13 (3): 342-356.