

CONTRÓLATE EN LAS REDES!

#BCN vs ODI

El discurso de odio
en las redes sociales:
un estado
de la cuestión



ajuntament.barcelona.cat/bcnvsodi

Ajuntament de
Barcelona



Área de Derechos de Ciudadanía, Cultura, Participación y Transparencia

Dirección de Servicios de Derechos de Ciudadanía y Diversidad

Barcelona, diciembre de 2017

Autoría del informe: Alex Cabo Isasi y Ana García Juanatey

TABLA DE CONTENIDOS

INTRODUCCIÓN	4
1.UN DIAGNÓSTICO DEL PROBLEMA	5
1.1.¿Qué se entiende por discurso de odio? Un concepto objeto de debate	6
1.2.Las particularidades del discurso del odio online	8
1.3.Las consecuencias del discurso del odio	9
1.4.¿Quiénes son los autores del discurso del odio?	9
2.LA RESTRICCIÓN DEL DISCURSO DEL ODIO	11
2.1.El dilema entre la libertad de expresión y la represión del discurso del odio	11
2.2.Las particularidades jurídicas del odio en Internet y las redes sociales	14
2.3.Legislación española	14
3.EL ROL DE LAS REDES SOCIALES COMO INTERMEDIARIAS	20
3.1.Atribución de responsabilidad	20
3.2.Auto-regulación e implementación de filtros informáticos	21
3.3.El Código de Conducta con la Unión Europea	23
4.LAS ESTRATEGIAS NO JURÍDICAS PARA COMBATIR EL DISCURSO DEL ODIO EN LAS REDES SOCIALES	25
4.1.Las estrategias de monitoreo e investigación	26
4.2.Las estrategias de presión a las redes sociales	27
4.3.Las estrategias orientadas a cambiar percepciones y actitudes	28
4.3.1. <i>Las campañas contra el discurso del odio en internet</i>	28
4.3.2. <i>Las estrategias contra el discurso del odio en Internet basadas en la educación y la formación</i>	31
CONCLUSIÓN	35
BIBLIOGRAFÍA	38

INTRODUCCIÓN

Se ha teorizado ampliamente sobre el potencial deliberativo y democratizador de Internet, y más en concreto, de las redes sociales. Su potencial participativo es innegable, y ha generado muchas esperanzas en las posibilidades de transformación política y social que ofrecen. Las redes sociales proporcionan un amplio acceso a fuentes de información ajenas al control de los gobiernos y las grandes corporaciones, facilitan la generación y coordinación de redes de activistas, y constituyen un espacio inmejorable para la deliberación y el intercambio de ideas. Las redes sociales, son, en palabras del experto en tecnologías de la información y la comunicación, Manuel Castells, una plataforma de “auto-comunicación de masas”, un espacio de autonomía comunicativa idóneo para que los actores sociales propicien la transición de la indignación a la esperanza. En este sentido, se considera la entonces denominada Primavera Árabe, el 15-M, u Occupy Wall Street como ejemplos de cómo las redes sociales pueden contribuir a transformar la indignación en acciones y movimientos horizontales, transformadores y emancipadores (Castells 2012).

Sin embargo, no conviene pecar de un exceso de optimismo, porque lo cierto es que internet y las redes sociales también pueden ser utilizadas con fines radicalmente diferentes a los expuestos. Tal y como afirma el ciberescéptico Evgeny Morozov, Internet no “*conduce necesariamente al respeto universal de los derechos humanos*”. Internet no es liberador ni democratizador en sí, sino que puede producir “*diferentes resultados políticos en diferentes entornos*”, y por tanto conviene no dejarse llevar por cierto “solucionismo tecnológico” o ciberutopismo (Morozov 2012).

De la misma manera que algunos movimientos emancipadores han hecho un esperanzador uso de las redes sociales, fuerzas de signo contrario han demostrado su capacidad de aprovechar ese potencial de las redes para fines opuestos. Gracias a Edward Snowden, se comprobó cómo utiliza Estados Unidos las nuevas tecnologías para la vigilancia y el espionaje indiscriminado. Otro ejemplo del lado oscuro de la Red es el denominado *50 Cent Army*, el “ejército” de comentaristas contratado por las autoridades chinas para circular mensajes propagandísticos y favorables al gobierno, en los debates en las redes.

Pero quizás el ejemplo más trágico, de cómo pueden ser utilizados internet y las redes, es el de la crisis desatada en Kenia tras las elecciones de diciembre de 2007. La propagación de mensajes incitadores a la violencia, a través de una variedad de foros en Internet, está documentada como “*un canal determinante a través del cual la violencia fue alimentada, llevándose las vidas de más de mil personas, y el desplazamiento de más de 600.000*” (Gagliardone et al. 2014: 5).

UN DIAGNÓSTICO DEL PROBLEMA

Efectivamente, este tipo de mensajes incitadores a la violencia ha encontrado en Internet y las redes sociales, el canal adecuado para su propagación. Aunque el alcance real del problema no ha sido determinado cuantitativamente, más que de una manera fragmentada y parcial, existe una sensación compartida por periodistas, juristas, ONGs, investigadores, y usuarios de las redes, en general, de que el problema del discurso extremo en las redes sociales es cada día más preocupante.

En toda Europa se está produciendo una gran proliferación de mensajes extremistas, asociada al contexto de la crisis de refugiados, con preocupantes “picos” de odio islamófobo detectados en las redes a raíz de los atentados de París, Bruselas, o Niza. Tras los atentados de París, *#matadatodoslosmusulmanes* se convirtió durante algunas horas en el tercer *hashtag* más utilizado en España (Jubany y Roiha 2016).

Según estudios del *think tank* Demos, en Twitter, hay aproximadamente unos 10.000 tuits por día con insultos racistas en lengua inglesa, lo que supone uno de cada 15.000 tuits (Gagliardone et al. 2014). En otro estudio realizado por Demos, sobre misoginia, que es otro de los motivos recurrentes de intolerancia, durante tres semanas de abril de 2016, entre usuarios británicos de Twitter, se encontraron más de 200.000 tuits con las palabras “zorra” o “puta”; constatándose que cada diez segundos, alguien insulta con esas palabras a una mujer en Twitter.

En el estado español el problema no es menor. A continuación, algunos ejemplos de 2016. Moha Gerehou, presidente de SOS Racismo Madrid, de origen gambiano, ha sido objeto de múltiples ataques. En julio de 2016, se le “subastaba” en Twitter como un esclavo o una pieza de caza, como respuesta a la campaña *#EstadoEspañolNoTanBlanco*. En septiembre del mismo año, fue amenazado de muerte tras colgar en Twitter, una foto de una concentración en contra de la retirada del callejero de Madrid del nombre de Millán Astray, fundador de la Legión¹.

En agosto de 2016, Jordi Ballart, alcalde de Terrassa recibió amenazas homófobas a través de Twitter y Facebook, a raíz de su decisión de retirar el nombre de una calle dedicada a un miembro de la División Azul².

En septiembre de 2016, la transexual Carla Antonelli, diputada de la Asamblea de Madrid, recibió amenazas de muerte a través de Twitter, por su defensa de los derechos del colectivo LGTB³. El mismo mes de septiembre, el árbitro de fútbol infantil, Jesús Tomillero, presidente de la asociación contra la LGTBfobia en el deporte, recibió insultos y amenazas de muerte en Twitter, por ser el primer colegiado auto-declarado gay en España⁴.

Estos son solo algunos casos recientes de acoso individual por motivos de odio e intolerancia, que han llegado a los medios de comunicación, y que constituyen la punta del iceberg de una

1. http://www.eldiario.es/sociedad/Nuevas-presidente-SOS-Racismo-Madrid_0_563094347.html

2. <http://www.elperiodico.com/es/noticias/sociedad/jordi-ballart-alcalde-terrassa-denuncia-amenazas-homofobas-redes-sociales-5473496>

3. http://ccaa.elpais.com/ccaa/2016/09/04/madrid/1472990074_481479.html

4. <http://www.elperiodico.com/es/noticias/deportes/arbitro-gay-jesus-tomillero-denuncia-amenazas-muerte-5376384>

situación que se ha convertido en habitual en las redes sociales. A nivel colectivo, el antigitanismo prolifera con impunidad en las redes, la islamofobia se convierte en tendencia de manera recurrente, a través de *hashtags* como #stopIslam, #terroristaswelcome o #musulmanesterroristas; el antisemitismo, la homofobia, la misoginia... Todas las ideologías intolerantes encuentran en las redes sociales un espacio de expresión privilegiado, que ha generado una especie de *cultura del odio*⁵, que contamina e intoxica las redes con lenguaje abusivo, denigrante o agresivo, por motivos, en gran medida, de intolerancia contra población inmigrada, refugiados, musulmanes, homosexuales, y otras minorías.

Las instituciones internacionales y europeas no son ajenas a este problema. Prueba de esta preocupación es la intensa actividad que ha desarrollado la Unión Europea durante el último año, en la búsqueda de soluciones para este problema; a pesar de que al mismo tiempo ciertas formas de racismo han ido ganando cada vez más terreno en el discurso político europeo. Como hitos destacados de esta dedicación destacan la firma del Código de Conducta con las empresas tecnológicas, y la puesta en marcha de un Grupo de Alto Nivel, en el que ha involucrado a representantes de los estados miembros de la Unión, del Parlamento Europeo, del Consejo de Europa, del Alto Comisionado de la ONU para los Refugiados (ACNUR), de la Organización para la Seguridad y la Cooperación en Europa (OSCE), de la Agencia Europea de Derechos Fundamentales (FRA), y de organizaciones de la sociedad civil como Amnistía Internacional, la Red Europea contra el Racismo (ENAR), la Plataforma Europea de ONGs sociales, y la Asociación Internacional de Lesbianas, Gays, Bisexuales e Intersex (ILGA), con el fin de prevenir y combatir los crímenes de odio y el discurso del odio, con una especial dedicación a la contra-argumentación del discurso del odio online.

También la UNESCO o la European Commission Against Racism and Intolerance (ECRI) han destacado, en informes recientes, el crecimiento del discurso intolerante en Internet y redes sociales, contra diferentes minorías, y la necesidad de encontrar estrategias eficaces para combatirlo.

En definitiva, el problema del discurso del odio en Internet, y en las redes sociales, en particular, es un asunto que ha adquirido tal dimensión que actualmente se encuentra en la agenda de un buen número de organismos europeos e internacionales. Las redes sociales se han convertido en un espacio de impunidad para la ira y el odio. El machismo, la homofobia, la xenofobia, la islamofobia, el antigitanismo, el antisemitismo y otras ideologías intolerantes aprovechan Internet y las redes sociales para el insulto, la humillación, el linchamiento social, el acoso, o la amenaza.

1.1. ¿Qué se entiende por discurso de odio? Un concepto objeto de debate

No hay una definición universalmente aceptada de discurso del odio. En primer lugar, porque el terreno de las definiciones con implicaciones éticas y jurídicas siempre es controvertido. Y en segundo lugar, porque el propio término “odio” lo convierte en un concepto emocional y abierto a la subjetividad. Es un concepto que genera confusión y que por su subjetividad, es relativamente manipulable.

El mínimo común denominador de cualquier definición de discurso del odio sería cualquier expresión de opinión o ideas basada en el desprecio y la animadversión hacia personas o colecti-

5. <http://time.com/4457110/internet-trolls/>

vos a los que se desea el mal. Sin embargo, esta definición simple englobaría un rango demasiado amplio de expresiones para que el concepto tuviera alguna utilidad para el análisis o la intervención social o jurídica (Article 19 2015).

Se habla de discurso del odio para referirse a las expresiones que incitan directamente a la comisión de actos de discriminación o violencia por motivos de odio racial, xenófobo, orientación sexual, u otras formas de intolerancia. Pero para algunos, el concepto se extiende también a aquellas expresiones que fomentan los prejuicios o la intolerancia, considerando que este tipo de expresiones contribuyen indirectamente a que se genere un clima de hostilidad que pueda propiciar, eventualmente, actos discriminatorios o ataques violentos (Gagliardone et al. 2015).

En el lenguaje cotidiano, y especialmente en los medios de comunicación, se ha generalizado el uso del concepto de discurso del odio para referirse a un conglomerado heterogéneo de manifestaciones que engloba desde amenazas a individuos o colectivos hasta casos en los que algunas personas simplemente expresan su ira contra las autoridades, de manera más o menos ofensiva (Gagliardone et al. 2015). Además se ha extendido en los medios de comunicación, e incluso en el ámbito jurídico, la utilización del término de discurso del odio para referirse a conceptos como la incitación o la apología del terrorismo, que aunque guardan cierta relación con el objeto de este informe, requieren un enfoque singularizado.

En definitiva, el término de discurso del odio es un concepto discutido, y que es objeto de debate político, jurídico y académico a nivel internacional. Es un concepto complejo, en la medida en que pone en juego y en colisión diferentes valores y principios capitales de los sistemas democráticos: la igualdad, la dignidad humana, la libertad de expresión, etc., cuya concepción no es idéntica en todos los contextos socio-políticos. Por ello, es un concepto, el del discurso del odio, que puede ser implementado políticamente con objetivos bien diferentes, que pueden ser más o menos legítimos.

Las definiciones que tienen mayor aceptación a nivel internacional se pueden agrupar en dos tendencias principales: 1) aquellas que definen el discurso del odio en un sentido más amplio y expansivo, e incluyen cualquier expresión que promueva o justifique el odio por motivos racistas, xenófobos, religiosos, de género, orientación sexual o discapacidad⁶; y 2) aquellas que lo definen de una manera más restrictiva y precisa, incluyendo sólo aquellas formas de expresión que, en contextos determinados de inestabilidad, puedan contribuir a desencadenar episodios violentos contra un grupo de personas por razón de su pertenencia a uno de los grupos señalados⁷ (Gagliardone et al. 2014).

La diferencia fundamental entre ambas se podría sintetizar en la dialéctica contenido-contexto, y en el mayor o menor riesgo de que los mensajes desencadenen acciones violentas. La primera de las tendencias conceptuales define el discurso del odio atendiendo exclusivamente a su contenido, mientras que la segunda considera que lo que convierte una determinada expresión en discurso del odio es el riesgo manifiesto de que, dado el contexto histórico-social en el que se emite, provoque episodios violentos. El contexto, efectivamente, determina en gran medida el daño que determinadas expresiones y manifestaciones pueden llegar a ocasionar, que puede

6. La Recomendación No. R(97)20 del Consejo de Europa, de 30/10/1997, lo definió como “toda forma de expresión que difunda, incite, promueva o justifique el odio racial, xenofobia, anti-semitismo u otras formas de odio basadas en la intolerancia (...)”

7. Susan Benesch (2013) propone el término “discurso peligroso (“dangerous speech”), para aquel discurso que tiene una probabilidad razonable de desencadenar episodios de violencia genocida.

oscilar desde el daño emocional hasta la provocación de escaladas de violencia masiva, como en los genocidios de Yugoslavia o Ruanda en los años 90.

La importancia de estas diferentes maneras de definir el discurso del odio estriba en que, aunque el concepto de discurso del odio no es una categoría estrictamente jurídica, estas definiciones informan las diferentes legislaciones y jurisprudencias nacionales, y las declaraciones, directivas, y convenios internacionales.

De cualquier manera, a los efectos de este informe, tomaremos como referencia las definiciones del primer grupo que conceptualizan el discurso del odio en un sentido relativamente amplio, porque recogen el tipo de expresiones que, dada su proliferación en el contexto europeo, merecen un rango más variado de intervenciones tanto de tipo jurídico como de tipo social.

1.2. Las particularidades del discurso del odio online

El discurso de odio online, para el que se ha acuñado la expresión de ciberodio, añade una serie de particularidades que lo convierten en un fenómeno descontrolado con un potencial de daño aún mayor. En primer lugar, la sobreabundancia comunicativa. Por medio de internet y las redes sociales, la comunicación de mensajes que antes estaban limitados al ámbito privado, se ha convertido en pública de manera absolutamente masiva. En segundo lugar, la descentralización de la comunicación, que “democratiza” las comunicaciones, en el sentido de cualquiera puede emitir un mensaje con un enorme potencial de audiencia. El efecto multiplicador de las redes sociales permite convertir un determinado mensaje en un fenómeno de transmisión exponencial, dando lugar a lo que se denomina metafóricamente como “viralizaciones”.

La permanencia de los contenidos, la itinerancia entre diferentes plataformas, el uso de pseudónimos, el anonimato, y la transnacionalidad son otras particularidades del espacio online que añaden desafíos en relación con el discurso del odio (Gagliardone et al. 2015). La permanencia y la itinerancia son particularidades que pueden aumentar el daño potencial del discurso del odio. El anonimato, el uso de pseudónimos y la transnacionalidad son características que dificultan su persecución, y que por la sensación de impunidad que ofrecen, alientan la expresión del discurso del odio.

En ocasiones, en el lenguaje coloquial nos referimos a Internet como un espacio “virtual”, adjetivo equívoco, en la medida en que se refiere a aquello “que tiene existencia aparente, pero no real”⁸. Esta sensación de “virtualidad”, como si lo que sucediera en Internet no fuera “real” y no tuviera repercusiones fuera de la Red, también actúa como efecto desinhibidor para la expresión de los discursos extremos.

En este sentido, Jubany y Roiha (2016) advierten de la falacia del dualismo digital, término concebido por Nathan Jurgenson para referirse a la falsa separación entre el mundo online y el mundo offline. La cuestión del discurso del odio en las redes sociales, aunque tenga las especificidades propias de esos espacios de comunicación, no es más que una exteriorización pública de un problema subyacente en la sociedad en general, independientemente del escenario en el que se manifieste.

8. Definición R.A.E.

1.3. Las consecuencias del discurso del odio

Las repercusiones que puede tener el discurso del odio, sea o no online, son variadas. Por un lado, hay que considerar el daño directo emocional o psicológico que pueden producir las amenazas, el acoso, y otros ataques dirigidos a individuos concretos, por motivos de odio e intolerancia. Por otro lado, el discurso del odio también produce un daño indirecto, erosionando la dignidad de las personas y dañando su reputación.

El discurso del odio juega un papel fundamental en la perpetuación de estereotipos discriminatorios, la estigmatización de colectivos, la contribución a su marginación, etc. Constituye un mecanismo lingüístico de importancia fundamental en la deshumanización de ciertos colectivos, a los que se trata de negar su pertenencia a la ciudadanía en condiciones de igualdad. En este sentido, el discurso del odio envía un mensaje que divide y segrega la sociedad. Pero al mismo tiempo, también juega un papel cohesionador para los emisores del discurso del odio, reforzando su sentimiento de pertenencia a un grupo (Gagliardone et al. 2015).

Mediante la proliferación y aceptación del discurso del odio se produce una reducción de empatía hacia los colectivos deshumanizados, y como consecuencia, se puede generar el caldo de cultivo adecuado para justificar actos discriminatorios, abusos y actos violentos de diversa naturaleza. Aunque no se pueda establecer, con carácter general, una conexión directa entre la proliferación de discurso del odio y los crímenes violentos de odio, cada vez hay una conciencia más clara del vínculo indirecto entre ambos fenómenos.

Si bien no todo discurso del odio desencadena episodios de violencia, lo que sí parece evidente es que raramente se producen episodios de crímenes de odio, sin una previa estigmatización y deshumanización de las víctimas. Por tanto, no está de más recordar aquí, los incidentes de delitos de odio registrados en 2015 por el Ministerio de Interior. Durante 2015, se denunciaron y registraron como delitos de odio 1328 casos, que incluyen lesiones, abusos sexuales, daños, tratos degradantes, etc. Movimiento contra la Intolerancia maneja estimaciones que sitúan la cifra real en torno a los 4000 casos. Teniendo en cuenta que el propio Ministerio de Interior reconoce que una gran parte de los delitos de odio no son denunciados, estas cifras son un indicio de hasta qué punto no es conveniente tomarse a la ligera el discurso del odio, como sustrato justificador o incitador de comportamientos violentos (Ministerio del Interior 2016).

1.4. ¿Quiénes son los autores del discurso del odio?

No cabe duda de que una gran parte del discurso del odio en las redes sociales procede de individuos directa o indirectamente vinculados con colectivos fascistas como Hogar Social Madrid, que cuenta con más de 15.000 seguidores en Twitter, grupos ultras simpatizantes de equipos de fútbol, o partidos políticos de ultra-derecha más o menos marginales como Plataforma per Catalunya, Falange Española de las JONS, España 2000, Alianza Nacional, etc.

Estos grupos han encontrado en las redes una útil herramienta de difusión de sus mensajes y su simbología fascista, que les sirve tanto para atacar a sus víctimas como para captar adeptos. Los individuos asociados a estos grupos, suelen actuar en las redes de manera colectiva, lo que provoca una mayor sensación de acoso a las víctimas de sus ataques. Además, para dar rienda suelta a sus incitaciones a la violencia en las redes, desarrollan diferentes códigos lingüísticos y

palabras clave, que les sirven para evitar tanto los filtros informáticos de las empresas tecnológicas, como la persecución judicial y policial de sus conductas⁹.

Sin embargo, lo cierto es que el fenómeno del discurso del odio no se agota en estos grupos con un marcado perfil ideológico. Hay otros perfiles de usuarios de las redes sociales, que en la jerga de internet son conocidos como *haters* o *troles*, que constituyen una parte importante del problema.

En Internet, se denomina *haters* a individuos que se dedican obsesivamente a atacar y agredir verbalmente a individuos concretos (en gran medida, personas famosas), o a colectivos a los que desprecian por su origen étnico, su religión, etc. Las redes sociales están plagadas de usuarios con una dedicación obsesiva al desprecio y la agresión verbal contra musulmanes, gitanos, homosexuales, mujeres, etc. Otro tipo de perfil asimilable sería el de aquellos usuarios con una tendencia a exteriorizar sus reivindicaciones y preocupaciones, sobre cualquier asunto, de manera extrema y agresiva, como una manera de atraer la atención y hacer más visibles sus mensajes.

El fenómeno del “troleo” también está ampliamente extendido en las redes, pero es muy arriesgado tratar de caracterizarlo como un fenómeno homogéneo. Algunos actúan a título individual para buscar atención o descargar su frustración. Otros actúan de manera colectiva, y se organizan para atacar a determinados objetivos. De cualquier manera, no cabe duda de que buena parte del fenómeno narcisista y sádico del “troleo” se produce con el simple ánimo de divertirse, provocando a otros usuarios de las redes sociales. De manera irreflexiva, y muchas veces ajenos al daño que pueden llegar a ocasionar, muchos usuarios de las redes se dedican a esparcir su hostilidad y agredir verbalmente a otros individuos o a colectivos minoritarios para divertirse con las reacciones que producen en los demás, bajo la coartada de un pretendido sentido del humor.

De cualquier manera, discernir a los que están intentando provocar para divertirse de los que están expresando sus verdaderas opiniones no siempre es fácil. Por ello, para abordar el fenómeno del discurso del odio en redes sociales, es preciso tener en cuenta la diversidad y la heterogeneidad de perfiles de sus autores, de cara, no sólo a dimensionar correctamente el problema, sino también para trazar estrategias adecuadas a los diferentes orígenes del discurso extremo. Con el uso de estos términos, propios de la jerga de internet se corre el riesgo de caricaturizar excesivamente el problema, y se produce la tendencia ya mencionada a considerarlo como algo ajeno a la realidad offline, que Nathan Jurgenson denomina falacia del dualismo digital. Las obsesiones hostiles de los *haters* o el sadismo narcisista de los *troles* tienen su espacio de expresión en las redes, pero sus motivaciones y sus prejuicios hostiles provienen del espacio offline.

9. <http://www.genbeta.com/redes-sociales-y-comunidades/asi-es-como-los-extremistas-camufan-su-discurso-de-odio-en-las-redes-sociales>

2. LA RESTRICCIÓN DEL DISCURSO DEL ODIO

2.1. El dilema entre la libertad de expresión y la represión del discurso del odio

Ante la creciente alarma generada, a nivel internacional, por la proliferación de discurso del odio, y las devastadoras consecuencias que tuvo en algunos genocidios como el de Ruanda o Yugoslavia, se han producido importantes movimientos y presiones para restringir el discurso del odio. Como resultado, los Estados han optado por diferentes niveles de restricción, cada uno en función de su contexto histórico y sociopolítico.

La diversidad de enfoques jurídicos de esta problemática, y el consiguiente nivel de represión del discurso del odio, tiene su origen en el importante debate que esta cuestión genera acerca de los límites de la libertad de expresión. La libertad de expresión, aunque es un pilar fundamental del sistema democrático, no es un derecho absoluto¹⁰. Todos los ordenamientos jurídicos contienen tipos penales que la restringen. Por ejemplo, no existe un derecho al insulto, la calumnia o la amenaza. En el caso del discurso del odio, la tensión se produce entre el derecho a la libertad de expresión, y el respeto a la dignidad humana y la igualdad. Mediante el discurso del odio, se busca la deshumanización y la exclusión social de minorías vulnerables, privándoles así de dignidad y del derecho a la igualdad.

En el ámbito internacional, las diferentes tendencias de restricción del discurso del odio se suelen sintetizar, de manera un tanto reduccionista, en la división entre la perspectiva estadounidense, más liberal y “tolerante”; y la perspectiva europea, más militante e “intransigente” con el discurso del odio.

- EE.UU. se ha caracterizado tradicionalmente por una defensa a ultranza de la Primera Enmienda¹¹, y por el “compromiso de sostener la libertad, incluso en circunstancias excepcionales” (Revenga-Sánchez, 2015). Se ha articulado la defensa de la libertad de expresión en una concepción del debate público como libre mercado de ideas, en el que la verdad siempre acaba obteniendo aceptación.
- Europa, y también otros países como Canadá, Australia, Sudáfrica, han seguido en general una línea más intransigente con el discurso del odio, y se han apoyado en la figura del abuso de derecho, para negar la posibilidad de invocar la libertad de expresión para amparar la negación del holocausto, la apología del terrorismo o mensajes abiertamente xenófobos o racistas.

Sin embargo, esta distinción entre modelos de entender los límites de la libertad de expresión no es tan simple. En primer lugar, la diferencia se ha visto atenuada en los últimos años, entre otros motivos, porque el tradicional respeto estadounidense por la libertad de expresión ha perdido

10. El Pacto Internacional de Derechos Civiles y Políticos de 1966 estableció límites a la libertad de expresión. El artículo 19 contempla la posibilidad de restricciones para “asegurar el respeto a los derechos o a la reputación de los demás” y para “la protección de la seguridad nacional, el orden público o la salud o la moral públicas”; y el artículo 20 prohíbe “toda apología del odio nacional, racial o religioso que constituya incitación a la discriminación, la hostilidad o la violencia”.

11. El Congreso no podrá hacer ninguna ley con respecto al establecimiento de la religión, (...) ni limitando la libertad de expresión, ni de prensa...”

algo de terreno como consecuencia del giro securitario producido en el marco de la denominada “guerra contra el terror”. Y en segundo lugar, porque tanto en Europa como en EE.UU. hay expertos y académicos que simpatizan y defienden la perspectiva y los argumentos contrarios a los que les correspondería por su tradición jurídica (Phillipson 2015).

Por otro lado, en el mundo musulmán, algunos países han presionado para que se adopten normas internacionales que prohíban la blasfemia, como una forma de odio religioso, en respuesta a casos como el de las caricaturas de Mahoma en el periódico danés Jyllands-Posten (Phillipson, 2015).

En el terreno académico, el debate sobre las prohibiciones del discurso del odio y la libertad de expresión es un debate que trasciende el ámbito jurídico y tiene implicaciones de orden político, ético y filosófico. Se ha enfocado el debate desde múltiples ángulos, y se plantean muy diversas cuestiones:

- ¿Son las prohibiciones del discurso del odio una amenaza para la libertad de expresión?
- ¿Se pueden considerar un mal necesario?
- ¿Prohibir el discurso del odio es un símbolo necesario del compromiso democrático con el respeto de la dignidad humana y la igualdad?
- ¿O es la normalización de las prohibiciones de discurso del odio síntoma de una democracia débil?
- ¿Son las prohibiciones del discurso del odio eficaces?

En el siguiente cuadro, se presentan de manera simplificada algunos de los principales argumentos que se suelen utilizar a favor y en contra de las prohibiciones del discurso del odio.

Tabla 1¹²

A favor	En contra
La libertad de expresión no es un derecho absoluto. Ha de ser limitado para evitar que se atente contra la dignidad humana, la igualdad, la paz y la convivencia social, el derecho a vivir sin acoso e intimidación, etc. El Pacto Internacional de Derechos Civiles y Políticos así lo reconoce en sus arts. 19.3 y 20.	La libertad de expresión y el debate público son esenciales para desarrollar una democracia plena. No deben establecer restricciones basándose únicamente en lo indeseables u ofensivas que puedan ser determinadas expresiones. Se corre el riesgo de abuso por parte de gobiernos que las utilicen como un “derecho penal del enemigo”, para castigar a la disidencia y los oponentes políticos.

12. Cuadro basado en Nineteen arguments for hate speech bans- and against them de Eric Heinze.

A favor

La proliferación de comentarios y expresiones ofensivas y extremas pueden generar un clima social que acabe desembocando en episodios de violencia y, en contextos conflictivos o sociedades divididas, en ataques genocidas.

El discurso del odio causa daños psicológicos directos (sentimiento de ser amenazado, humillado, etc.) y daños indirectos, en el sentido de que contribuye a la perpetuación de situaciones de discriminación.

Las declaraciones más extremas de odio son incitadoras de actos violentos y de discriminación, y por ello deben ser restringidas y castigadas.

El estado no puede mantenerse neutral, y las prohibiciones del discurso del odio son un símbolo y una declaración de principios del compromiso con los valores democráticos, la igualdad, y la dignidad humana.

En contra (*continuación*)

En democracias consolidadas y prósperas, no se ha demostrado que exista una correlación clara entre proliferación de discurso del odio y un aumento de crímenes de odio. El estado ya puede castigar la motivación de “odio” de los actos criminales como circunstancia agravante.

Existen medios legales para prohibir el acoso, las amenazas y los ataques individuales a la dignidad; e instrumentos más legítimos, proporcionados y eficaces para combatir la discriminación: leyes de no discriminación laboral, educación en la diversidad y el pluralismo, campañas de concienciación, etc.

El estado ya puede castigar actos de provocación o conspiración, en los que se puede establecer una clara conexión material con la comisión de un delito; pero el concepto de incitación le otorga un mecanismo para castigar sin tener que demostrar la posibilidad de que se realice el daño como consecuencia de la expresión pública de ideas.

El estado tiene vías más legítimas y eficaces para posicionarse a nivel simbólico. Las prohibiciones son poco eficaces, e incluso contraproducentes, porque se corre el riesgo de deslegitimar la promoción real de la igualdad y porque pueden convertir a ciertos individuos en “mártires” de la libertad de expresión.

En relación con la eficacia o al carácter contraproducente de las prohibiciones, es muy interesante el estudio realizado en Holanda por Van Spanje, J. y de Vreese, C. (2015), sobre el impacto que la judicialización del discurso del odio de determinados políticos de extrema derecha tiene sobre el apoyo a sus partidos. Estos investigadores encontraron evidencias empíricas de que el inicio de una causa judicial contra el discurso anti-inmigración de Geert Wilders, reforzó el atractivo de su partido para los votantes, y sugieren que la decisión de judicializar sus declaraciones islamófobas contribuyó a su despegue electoral (Van Spanje y de Vreese 2015).

En relación con la posibilidad de que se produzcan abusos políticos del concepto de discurso del odio, un caso paradigmático es de la Sudáfrica del Apartheid, donde se utilizaban las leyes contra este tipo de discurso para criminalizar las críticas contra la dominación blanca (Gagliardone et al, 2015). También en el estado español, hemos presenciado recientemente diversos ejemplos de uso del concepto de discurso del odio contra la disidencia y los enemigos políticos,

por ejemplo, el caso de César Strawberry, líder de Def-Con-Dos¹³, al que el fiscal la Audiencia Nacional acusaba de incitar al odio; el de Guillermo Zapata, concejal del Ayuntamiento de Madrid, al que se encausó por unos chistes publicados en Twitter cuando debatía sobre los límites del humor negro¹⁴; o el caso de la joven valenciana María Lluch, “Madame Guillotine” en Twitter, que fue juzgada por hacer chistes sobre víctimas de ETA, y condenada a 1 año de prisión por considerar el Tribunal Supremo sus tuits como discurso del odio no protegido por la libertad de expresión¹⁵, pese a rebajar la condena inicial de 2 años, por desproporcionalidad. Estos casos, independientemente del juicio ético que a cada uno le puedan merecer, ilustran la problemática de cómo el concepto de discurso del odio puede utilizarse para justificar el castigo penal de ciertas manifestaciones proferidas por oponentes políticos.

No se puede perder de vista, por tanto, la idea de que cualquier legislación, incluida la relativa a la restricción del discurso del odio es la expresión del grupo dominante que controla el contenido de la ley (Gagliardone et al. 2015).

2.2. Las particularidades jurídicas del odio en Internet y las redes sociales

Por si este debate político, ético y jurídico no tuviera suficientes aristas, Internet y las redes sociales añaden aun mayor complejidad al problema del discurso del odio. En primer lugar, porque Internet se ha convertido en una especie de fetiche de liberación, considerado utópicamente por muchos como *“una fuerza revolucionaria que no debería estar sometida a regulación alguna”*, y cualquier intento de regular lo que sucede en la Red, es automáticamente tachado de reaccionario, ilegítimo y antidemocrático (Morozov, E., 2012).

Internet y las redes sociales plantean un segundo problema, consustancial al carácter “sin fronteras” de la Red: el problema de la delimitación de la cuestión jurisdiccional. Un mensaje emitido haciendo uso de un servidor de EE.UU. puede tener consecuencias lesivas en España. ¿Qué normativa es aplicable? ¿La estadounidense, más tolerante, o la española?

Esta extraterritorialidad supone también un desafío en materia de cooperación judicial, en la medida en que las diferencias legislativas también afectan a materias de importancia capital en el ámbito de la cibercriminalidad, como la protección de datos o el secreto de las comunicaciones.

Se plantean, además, problemas derivados de la propia configuración técnica y la operatividad de Internet (localización de servidores, autenticación de IPs, cuentas robot, diferentes procedimientos de encriptación para ocultar la identidad de los ataques, etc.) que plantean otra serie de incertidumbres y dificultades en cuestión de obtención de pruebas o de determinación de responsabilidad.

Sin entrar aquí en un terreno de gran complejidad jurídica, la transformación del modelo comunicativo que supone Internet sitúa el debate relativo al discurso del odio ante dos alternativas. La primera, la búsqueda de un diálogo internacional para lograr un *“acuerdo global sobre los criterios de restricción o permisividad del discurso del odio en internet”*, que se antoja poco realista. Y la segunda, la *“imposición a Internet de las políticas constitucionales de cada estado”* (Rodríguez)

13. http://www.eldiario.es/politica/Absuelto-Cesar-Strawberry-enaltecido-Twitter_0_538946313.html

14. <http://www.lavanguardia.com/politica/20160307/40269028704/archivada-causa-zapata-tercera-vez-tuits.html>

15. <http://www.poderjudicial.es/cgpj/es/Poder-Judicial/Noticias-Judiciales/El-Tribunal-Supremo-condena-a-un-ano-prision-a-una-joven-por-humillar-a-traves-de-twitter-a-Irene-Villa-y-a-Miguel-Angel-Blanco>

guez Izquierdo, 2015), que presenta los límites e incertidumbres jurídicas que se han expuesto aquí superficialmente.

2.3. Legislación española

La legislación española relativa al discurso del odio en las redes sociales está condicionada por la necesidad de ajustarse a las normas europeas, que siguen la tendencia, cada vez más generalizada a nivel global, de introducir restricciones penales al discurso del odio¹⁶.

La última reforma del Código Penal¹⁷ responde a la exigencia de ajustarse a dicha regulación europea, y a la sentencia del TC 235/2007, de 7 de noviembre, que impuso una interpretación restrictiva del delito de negación del genocidio, limitando su aplicación a los casos en que esta conducta implique una incitación al odio u hostilidad contra minorías.

La nueva regulación ha unificado los diferentes delitos de discurso del odio en el artículo 510 del Código Penal, y los castiga con las siguientes penas:

1. 1 a 4 años de prisión y multa de 6 a 12 meses,
 - a. Incitación directa o indirecta **al odio, a la hostilidad, a la discriminación y a la violencia**, contra grupos o individuos por motivos racistas, antisemitas u otros relativos a su ideología, religión, origen, sexo, orientación sexual, enfermedad o discapacidad.
 - b. a. Producción y difusión de **materiales**, con la misma motivación que en el apartado a).
 - c. **Negación o enaltecimiento** de los delitos de **genocidio**, lesa humanidad o contra las personas o bienes protegidos en caso de conflicto armado que hubieran sido cometidos contra los grupos señalados, cuando ello promueva o favorezca un **clima de violencia, hostilidad u odio** contra los mismos.
2. penas de prisión de 6 meses a 2 años y multa de 6 a 12 meses (y prisión de 1 a 4 años cuando se promueva un clima de violencia, odio o discriminación)
 - a. Realización de actos, o producción o difusión de materiales, que supongan **humillación o menosprecio** contra alguno de los grupos señalados o de sus miembros por motivos discriminatorios;
 - b. el **enaltecimiento o justificación** pública de los **delitos** cometidos contra los grupos señalados.

La regulación del art. 510 establece además una serie de agravantes y consecuencias accesorias. A los efectos de este informe, es destacable la agravante por haberse llevado a cabo, cual-

16. En concreto, la Decisión Marco 2008/913/JAI, de 28 de noviembre, relativa a la lucha contra determinadas formas y manifestaciones de racismo y xenofobia mediante el Derecho penal, de la Unión Europea, establece unas directrices básicas que deben recogerse en todos los estados miembros de la Unión. También es destacable en este sentido, el Protocolo Adicional al Convenio sobre la ciberdelincuencia, relativo a la penalización de actos de índole racista y xenófoba cometidos por medio de sistemas informáticos de 2003, del Consejo de Europa.

17. L.O. 1/2015, de 30 de marzo.

quiera de las conductas descritas, haciendo uso de las “nuevas tecnologías”¹⁸. En ese caso, atendiendo al potencial mayor alcance de las conductas punibles, las penas se impondrán en su mitad superior, y se prevé la retirada de los contenidos, y el bloqueo del acceso o la interrupción de la prestación del servicio¹⁹.

Para parte de la doctrina, la reforma viene a corregir algunas de las deficiencias de la regulación anterior. La principal aportación es que elimina la palabra “provocación” que había generado dificultades de aplicación del artículo, disparidad de criterios y sentencias contradictorias, al interpretarse en el sentido del artículo 18 del Código Penal, de “incitación directa”. Al eliminarse ese término de “provocación”, se facilita que se castigue también la “incitación indirecta”, que está además recogida explícitamente en el texto (Gascón Cuenca 2015).

También considera, esta parte de la doctrina, que la inclusión, en el apartado 2 del artículo, de dos delitos atenuados para aquellas conductas *“que no reúnen todas las características como para incitar a la hostilidad, a la discriminación o a la violencia, pero que sí son ideales para lesionar la dignidad de los individuos”*, con penas inferiores (6 meses a 2 años de cárcel), adecua mejor la regulación al principio de proporcionalidad (Gascón Cuenca 2015).

Sin embargo, para otros autores, esta reforma es técnicamente pobre y reaccionaria, y sigue planteando dudas de su constitucionalidad, por falta de adecuación a la justificación de la intervención del derecho penal, y por seguir adoleciendo de falta de proporcionalidad. Algunos penalistas consideran incluso que se ha profundizado en la mayoría de los defectos de la regulación anterior, dejando demasiado a expensas del criterio de los jueces el sancionar o no conductas que aparentemente encajan en los tipos regulados por el art. 510 (Teruel Lozano 2015).

Para los críticos con la nueva regulación, existen mejores y más proporcionadas respuestas que la intervención penal. El ordenamiento español cuenta con *“instrumentos suficientes para castigar”*, por ejemplo las humillaciones, las amenazas o la provocación de un peligro cierto e inminente de actos ilícitos, *“sin que sea necesario plantear una respuesta excepcional a través de tipos penales ad hoc anti-discriminatorios”* (Teruel Lozano 2015: 39).

Según Rey Martínez (2015: 52), los problemas parten de que se ha realizado un enfoque de la cuestión demasiado inclinado hacia el derecho penal. Para él, *“una sanción administrativa e, incluso, la exigencia de responsabilidad civil, podrían ser mucho más respetuosas”* con la Constitución y además, serían probablemente más eficaces que una norma penal, que por su rigurosidad y desproporcionalidad, puede acabar resultando inaplicable para los jueces.

Efectivamente, hasta la fecha, la jurisprudencia ha hecho una interpretación restrictiva de los delitos de discurso de odio. Esto ha tenido como resultado un panorama jurídico de penas muy duras y nula sanción del discurso del odio. Para Rey Martínez (2015: 77), se precisa *“más sanción de los discursos del odio, pero con una dureza penal proporcionada”*. La nueva regulación, a pesar de la introducción de los tipos atenuados del apartado 2 del artículo 510, ha mejorado poco en este aspecto, al establecer penas de hasta 4 años de cárcel, apartándose incluso de la Decisión Marco 2008/913/JAI de la UE, que estipula penas máximas de 3 años (Gascón Cuenca, 2015).

18.Art. 510.3 CP

19.Art. 510.6 CP

Veremos cuál es la interpretación que los jueces y magistrados hacen de la nueva regulación. En cualquier caso, sería deseable que no incurrieran en las contradicciones que se produjeron al aplicar la legislación anterior, y que algunos expertos, como Miguel Ángel Aguilar, fiscal coordinador de Delitos de Odio de la Fiscalía de Barcelona, criticaron duramente²⁰, al dictar el Tribunal Supremo dos sentencias contradictorias en el plazo de una semana.

De cualquier manera, no se ha de olvidar que la represión penal de lo que genéricamente se entiende por discurso del odio no se agota en el art. 510 del Código Penal. El Código Penal contempla una circunstancia agravante genérica por motivos discriminatorios, que agrava la responsabilidad penal de cualquier delito, y que, por ejemplo, puede aplicarse a conductas típicas en las redes sociales²¹, como las amenazas o el acoso. La consideración de que concurre la agravante de discriminación supone que la pena correspondiente a cada delito se impondrá en su mitad superior. Además existe un delito de amenazas a colectivos,²² que castiga las amenazas, con la entidad suficiente como para lograr atemorizar, dirigidas a grupos étnicos, culturales, etc.

Principales problemas para la persecución de los delitos de discurso del odio

Algunos de los principales problemas, de orden práctico, que dificultan una efectiva persecución de los delitos de odio en general, y de los delitos de discurso del odio en redes sociales, en particular, son, a juicio de los expertos (Aguilar et al., 2015), los siguientes:

- Desconocimiento de los hechos por el escaso número de denuncias.
- Minimización de los hechos en algunos casos por parte de los responsables de las investigaciones.
- Insuficientes investigaciones policiales y judiciales.
- Dificultades para acreditar la motivación “odiosa” o discriminatoria del delito.
- Poca formación en igualdad y no discriminación de las instituciones implicadas: cuerpo judicial, funcionarios judiciales, forenses, policía, seguridad privada, etc.
- Dificultades inherentes al escenario “redes sociales”: problemas de jurisdicción, territorialidad, identificación y obtención de datos de los agresores, etc.

Normas administrativas

Desde el punto de vista del derecho administrativo, existen algunas normativas que contemplan la posibilidad de sancionar conductas que pueden ser entendidas como discurso del odio. En este sentido, la normativa más específica es la Ley 19/2007, contra la violencia, el racismo, la xenofobia y la intolerancia en el deporte, que es un ámbito de especial incidencia de conductas intolerantes, y en el que los grupos ultras de extrema derecha encontraron durante años un es-

20.<http://www.elmundo.es/elmundo/2011/06/13/barcelona/1307982347.html>

21.Art. 22.4 CP

22.Art. 170.1 CP

pacio propicio para dar rienda suelta a sus actitudes abiertamente racistas, homófobas, y discriminatorias, con relativa impunidad.

En lo que respecta al objeto específico de este informe, esta ley considera la difusión, por medios informáticos o tecnológicos, vinculados a información deportiva, “de contenidos que promuevan o den soporte a la violencia, o que inciten, fomenten o ayuden a los comportamientos violentos o terroristas, racistas, xenófobos o intolerantes por razones de religión, ideología, orientación sexual (...)” como infracción muy grave, y la castiga con importantes sanciones económicas.²³

Instituciones especializadas y protocolos policiales contra el odio y la discriminación

En marzo de 2013, el Ministerio Fiscal puso en marcha una red de **fiscales especializados contra el odio**, con la designación de un fiscal en cada provincia, encargado de coordinar la actuación, en materia de odio y discriminación, de la Fiscalía de la provincia, y de coordinar dicha actuación con el resto de la red de fiscales provinciales especializados en la materia.

Esta figura se inspiró en el ejemplo del Servicio de Delitos de Odio y Discriminación de la Fiscalía de Barcelona, constituido en 2009, y que fue pionero en este ámbito, y ha sido reconocido por la Agencia de Derechos Fundamentales de la Unión Europea, como un ejemplo de buena práctica, que debería ser reproducido en otros estados de la UE.

Entre los retos que se plantea este servicio, se encuentran los siguientes (Aguilar et al. 2015):

- recabar mejor información sobre hechos que pudieran constituir conductas punibles en esta materia;
- promover la comunicación con la Fiscalía de los colectivos y ciudadanos afectados, tanto para denunciar como para compartir iniciativas y sugerencias;
- contribuir a unificar los criterios de aplicación e interpretación del Código Penal
- afrontar la complejidad de las investigaciones de los delitos de incitación al odio, la violencia y la discriminación, cuando se producen en Internet y redes sociales;
- recopilar recomendaciones y jurisprudencia emitida por organismos y tribunales internacionales en materia de no discriminación;
- combatir a los grupos violentos organizados que difunden el discurso del odio;
- afrontar la tendencia a minimizar la importancia de los delitos de odio, por algunos policías, jueces y fiscales;
- mejorar la coordinación con las Fuerzas y Cuerpos de Seguridad del Estado.

En este sentido, también es reseñable el desarrollo de **protocolos de intervención policial**. Los Mossos d'Esquadra fueron los pioneros en esta materia al dotarse de un “Procedimiento de

23.Art. 23, 1, c) y 24, 1, c) de la Ley 19/2007. Sanciones desde 60.000,01€ a 650.000€.

hechos delictivos motivados por el odio o la discriminación” en 2010. El Ministerio del Interior publicó en 2014 el “Protocolo de actuación para las Fuerzas y Cuerpos de Seguridad para los delitos de odio y conductas que vulneran las normas legales sobre discriminación”, dirigido a la Policía Nacional y la Guardia Civil, basándose en materiales editados por el Observatorio Español del Racismo y la Xenofobia, del Ministerio de Empleo y Seguridad Social (Aguilar García 2015).

Estos protocolos responden a la necesidad de formar a los cuerpos policiales y dotarles de unas pautas en la recogida de datos estadísticos, la elaboración de los atestados, la investigación de los hechos para acreditar la existencia de una motivación racista o intolerante, la atención y la protección de las víctimas, o la interlocución con la comunidad y las ONGs que representan a las víctimas.

El Protocolo del Ministerio del Interior contempla recomendaciones específicas para los delitos de odio cometidos a través de Internet y las redes sociales. En concreto, y dadas las dificultades que el uso de las tecnologías de la información supone para la persecución e identificación de los responsables de las infracciones, recomienda que las investigaciones sean realizadas por Unidades especializadas de la Policía Judicial (Ministerio de Interior 2015).

Además, tanto las instituciones públicas como algunas organizaciones no gubernamentales (SOS Racismo, Movimiento contra la Intolerancia, Observatorio contra la Homofobia, Fundación Rais, etc.) ofrecen una serie de servicios orientados a asesorar y asistir jurídicamente a las víctimas de los delitos de odio, en general, y del discurso del odio, en particular. En Catalunya, los Ayuntamientos de Sabadell, Terrassa, y Barcelona (Oficina per la No Discriminació, OND) cuentan con diferentes servicios, que ofrecen asesoramiento jurídico en este campo.

3. EL ROL DE LAS REDES SOCIALES COMO INTERMEDIARIAS

Indiscutiblemente, las propias redes sociales, como canal de comunicación de los mensajes “odiosos”, juegan un papel crucial en la lucha contra el discurso del odio. Basta recordar que la red social líder, Facebook, cuenta con más de 1.700 millones de usuarios activos²⁴, que sería aproximadamente el equivalente a la suma de la población de los continentes europeo y americano. Estas cifras masivas convierten a las redes sociales en un actor, con un rol de intermediario, pero con un poder innegable.

Su papel de intermediarias en la comunicación digital, les convierte en el primer árbitro en la determinación de qué se puede decir y qué no. En este sentido, conviene recordar que las grandes redes sociales más usadas en Europa, tienen su sede en Estados Unidos, y por tanto, están imbuidas de la filosofía norteamericana, más defensora de una libertad de expresión más absoluta y más tolerante con cierto discurso del odio. Esta circunstancia tiene, como hemos señalado en el epígrafe anterior, repercusiones legales, por ejemplo, en términos de colaboración con la justicia en cuanto a requerimientos de datos. Pero también influye en sus políticas de auto-regulación y en la forma de llevarlas a la práctica.

En relación con las empresas intermediarias, se plantean diferentes alternativas:

- la atribución de responsabilidad a las empresas,
- la promoción de la autorregulación,
- y la implementación de filtros informáticos (Rodríguez Izquierdo 2015).

Cualquiera de estas vías plantea ciertos dilemas.

3.1. Atribución de responsabilidad

Mientras que en Estados Unidos los proveedores de servicios están prácticamente exentos de responsabilidad por los contenidos publicados, la atribución de responsabilidad a los prestadores de servicios intermediarios en la Unión Europea está basada principalmente en el conocimiento efectivo del carácter ilícito de los contenidos publicados en su servicio (Rodríguez Izquierdo 2015). En principio, sólo serán responsables cuando no retiraran o impidieran el acceso a un contenido ilícito con prontitud, habiendo tenido conocimiento de su ilicitud por medio de una notificación del titular de los derechos o por medio de una orden de un órgano competente.

La sentencia Delfi del Tribunal Europeo de Derechos Humanos estableció la responsabilidad del portal de noticias estonio, Delfi, al considerarlo como proveedor de contenidos, por los comentarios publicados por sus lectores, incluso sin haber recibido una solicitud de retirada del contenido ilícito. Sin embargo, en la sentencia, el Tribunal determinó que esta interpretación no es aplicable a las redes sociales, estableciendo una diferenciación entre proveedores de servicios

²⁴. <http://newsroom.fb.com/company-info/>

técnicos, con un rol pasivo, como las redes sociales, y los proveedores de servicios de contenidos, como una web de noticias o un medio de comunicación online²⁵.

3.2. Autoregulació i implementació de filtres informàtics

La persecución y restricción de los delitos de discurso del odio por los estados se enfrenta a las dificultades ya señaladas para actuar fuera de sus fronteras. Sin embargo, y aquí reside una de las principales ventajas de las estrategias de auto-regulación, las compañías, en tanto que operadores privados, tienen el derecho de establecer códigos éticos y condiciones de uso de sus servicios, aplicables globalmente.

Todas las grandes redes sociales, Twitter, Facebook, Instagram, Tumblr, etc., e incluso los medios de comunicación, cuyos foros de comentarios también son un canal habitual del discurso del odio, cuentan con políticas o condiciones de uso que, en principio, prohíben el discurso del odio en sus plataformas de comunicación.

Twitter, por ejemplo, prohíbe expresamente²⁶ las amenazas, la incitación a la violencia, el acoso, así como el comportamiento que incita al odio, por motivos de origen étnico, nacionalidad, orientación sexual, género, religión, etc.

Facebook incluye en sus condiciones de servicio cláusulas relativas a la no publicación de contenido intimidatorio, discriminatorio, incitador de la violencia, pero reconoce expresamente la incapacidad de garantizar que Facebook sea un lugar seguro, y enuncian sus normas a modo de compromisos del usuario. Desde la perspectiva de que Facebook constituye una comunidad global, sometida a diferentes legislaciones nacionales, establecen unas normas comunitarias a modo de auto-regulación.²⁷ Facebook expone, mediante dichas normas, qué contenido se puede denunciar y eliminar, destacando, a los efectos de este informe, la eliminación del lenguaje que incita al odio por motivos de raza, etnia, nacionalidad, religión, orientación sexual, sexo, género, identidad sexual, o discapacidades o enfermedades graves.

La mayoría de las redes sociales, incluidas las dos grandes redes sociales, Facebook y Twitter, cuentan con sistemas de denuncia que permiten a los usuarios poner en conocimiento de las empresas la presencia de contenido abusivo que vulnere las normas de uso de las plataformas, para que las compañías valoren su eliminación. Facebook advierte expresamente en sus normas comunitarias, de que no todo contenido ofensivo infringe sus normas, y por ello, ofrece además la posibilidad de bloqueos y filtros personalizados para ocultar aquel contenido que el usuario no desea ver.

Sin embargo, a pesar de esta actitud, teóricamente intolerante con el discurso del odio, por parte de las compañías, lo cierto es que los mecanismos de denuncia y eliminación de comentarios se llevan a la práctica de manera un tanto disfuncional. Así lo atestigua una investigación del proyecto PRISM (Jubany y Roiha 2016), en la que de un total de 100 denuncias realizadas a Facebook, solo 9 obtuvieron como resultado la eliminación de los comenta-

25. <http://blog.garrigues.com/se-confirma-la-responsabilidad-de-los-portales-de-noticias-por-los-comentarios-difamatorios-de-los-lectores/>

26. <https://support.twitter.com/articles/72688>

27. <https://www.facebook.com/communitystandards#>

rios²⁸. La errática política de eliminación de comentarios de la compañía, ya sea por el mal funcionamiento de los algoritmos que emplean o porque los equipos encargados de gestionar las denuncias estén superados por el ingente volumen de comentarios denunciados, tuvo como resultado la eliminación de mensajes como “Los perros son más discretos que los musulmanes”, mientras que mensajes como “Matarlos a todos sin piedad” fue considerado por Facebook, como un mensaje que no vulneraba las normas de comunidad.

Una de las principales críticas que se han venido haciendo sobre los mecanismos de denuncia de las redes sociales es su falta de transparencia. No se sabe a ciencia cierta si el proceso lo realizan mediante algoritmos, mediante equipos humanos o con una mezcla de ambos. Por otra parte, tampoco hacen públicos datos sobre el número de denuncias, los motivos de las denuncias, o el porcentaje de denuncias que finalizan con comentarios o cuentas eliminadas (Jubany y Roiha 2016).

Facebook ha desarrollado también 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. El caso de Shaun King, periodista de New York Daily News y activista de Black Lives Matter, fue especialmente sintomático. A pesar de que había sido invitado meses antes por Facebook como ponente para hablar de este movimiento político contra la violencia hacia las personas negras, su cuenta de Facebook, fue suspendida por publicar un email en el que le dirigían insultos racistas. Gracias a sus contactos en Facebook, el activista logró que se le reestableciera el servicio, pero el propio activista se cuestionaba cómo puede gestionar un caso similar una persona que no tenga el acceso privilegiado a determinadas instancias de la compañía²⁹.

Las propias redes sociales son conscientes del problema que tienen con la gestión del acoso online, las amenazas y el discurso del odio. En 2015, se filtró un comunicado interno del entonces CEO de Twitter, Dick Costolo, en el que reconocía que tenían un grave problema en la gestión de este asunto, que les estaba haciendo perder usuarios cada día y limitando en sus objetivos de crecimiento³⁰. Por esa razón, Twitter ha empezado a considerar prioritario este asunto, y está inmerso en un proceso de cambios que permitirán a los usuarios un control más proactivo de su cuenta de Twitter. La compañía ha buscado el asesoramiento de una comisión externa de organizaciones de la sociedad civil involucradas en la prevención de abusos online³¹, y están trabajando, entre otras herramientas, en la puesta en marcha de filtros de calidad para bloquear determinadas palabras, *hashtags*, insultos racistas, etc. para que no aparezcan en su *timeline*. Su objetivo es encontrar mejores mecanismos para bloquear a otros usuarios, y agilizar y optimizar los sistemas de denuncia a la compañía de comportamientos abusivos.

Un caso paradigmático, en esta línea de menor tolerancia de Twitter con el odio online, se produjo en julio de 2016, con la suspensión temporal de Milo Yiannopoulos, un icono del movi-

28. <http://www.elperiodico.com/es/noticias/sociedad/estudio-proyecto-prism-denuncia-banalizacion-discurso-odio-internet-5284591>

29. http://www.eldiario.es/theguardian/Facebook-temporalmente-Black-Lives-Matter_0_558544311.html

30. <http://www.lavanguardia.com/tecnologia/20160901/4122751789/twitter-acoso-trolls-redes-sociales.html>

31. <https://about.twitter.com/es/safety/council>

miento *alt-right* en Estados Unidos, por liderar supuestamente el acoso misógino y racista a la actriz afroamericana Leslie Jones. El carácter mediático de Milo, y el hecho de que contara con 300.000 seguidores en Twitter, convirtió el caso en un debate sobre la libertad de expresión, convirtiéndose en tendencia el *hashtag* #FreeMilo, que lo convertía en una especie de mártir de la libertad de expresión³².

No cabe duda de que el hecho de que la atacada fuera una celebridad de Hollywood, facilitó la visibilidad y la resolución del caso, en el que intervino públicamente el actual CEO de la compañía, Jack Dorsey. El problema, más allá de la importancia de casos simbólicos como este, lo tienen las personas “anónimas” que sufren humillaciones y vejaciones en las redes sociales, pero no tienen la capacidad de que sus casos lleguen a tan altas instancias. De momento, habrá que esperar para comprobar si los cambios tecnológicos en los que trabaja la compañía, suponen un cambio real en la gestión del problema del abuso y el odio online.

De cualquier manera, para equilibrar y matizar las críticas que se suelen hacer a estas compañías, es justo recordar aquí la enorme dificultad de la tarea a la que se enfrentan. Twitter, por ejemplo, maneja unos 300 millones de tuits al día, y han llegado a picos de 600 millones³³. Incluso con la ayuda de filtros y algoritmos informáticos, la tarea de moderación y autorregulación es de proporciones gigantescas, y al mismo tiempo, muy delicada, en la medida en que lidian con conceptos tan sensibles para su imagen, como la libertad de expresión o la censura.

3.3. El Código de Conducta con la Unión Europea

Como consecuencia de la crisis de refugiados de 2015, y ante el evidente crecimiento de las manifestaciones xenófobas y racistas en las redes sociales, la Unión Europea presionó a las empresas tecnológicas para que asumieran un rol más activo en la lucha contra el discurso del odio.

Como resultado de esa presión de algunos estados y de la propia Unión Europea, en mayo de 2016, Facebook, Twitter, Youtube y Microsoft firmaron un Código de Conducta en materia de incitación ilegal al odio en internet³⁴. En dicho acuerdo, las empresas tecnológicas adquirieron una serie de compromisos, entre los que destaca el compromiso de revisar las solicitudes de retirada de contenidos ilegales de incitación al odio en el plazo de 24 horas. Además, se comprometieron a establecer procedimientos claros de examen de los contenidos denunciados, y a valorar las denuncias que reciban en sus servicios, con arreglo a sus normas de autorregulación, pero teniendo en cuenta también, cuando sea necesario, las legislaciones nacionales de transposición de la Decisión Marco europea relativa a la lucha contra determinadas formas y manifestaciones de racismo y xenofobia mediante el Derecho penal, que el acuerdo considera su base jurídica.

El acuerdo busca la cooperación entre las empresas firmantes y el intercambio de buenas prácticas con otros servicios de comunicación social de Internet, cuya adhesión al acuerdo se promoverá.

32.<http://www.bbc.com/mundo/noticias-36847433>

33.<http://uk.businessinsider.com/tweets-on-twitter-is-in-serious-decline-2016-2>

34.http://europa.eu/rapid/press-release_IP-16-1937_es.htm

Las empresas firmantes también se comprometen a promover iniciativas que desarrollen un “contradiscursio” alternativo, a apoyar programas educativos que fomenten el pensamiento crítico, y a colaborar con las organizaciones de la sociedad civil en actividades formativas. Este compromiso firmado en el Código de Conducta formaliza y compromete una línea de actuación que ya venían desarrollando las empresas tecnológicas en este ámbito para mejorar la manera de lidiar con el acoso y el odio en las redes.

El cumplimiento del código de conducta y sus resultados serán sometidos a evaluación por el recientemente creado **Grupo de Alto Nivel sobre la Lucha contra el Racismo, la Xenofobia, y otras formas de Intolerancia**, que está formado por los estados miembros de la UE, representantes del Parlamento Europeo, del Consejo de Europa, del Alto Comisionado de la ONU para los Refugiados, de la Organización para la Seguridad y la Cooperación en Europa, de la Agencia Europea de Derechos Fundamentales, y de organizaciones de la sociedad civil.

Aunque la Comisaria europea de la Dirección General de Justicia, Consumidores e Igualdad de Género, Vera Jourová, presentó este acuerdo como un punto de inflexión, a partir del cual cambiarían las reglas del juego³⁵, lo cierto es que la firma de este código de conducta no ha estado exenta de críticas. En primer lugar, se ha lamentado que el proceso de negociación haya estado exento de transparencia y que no se hayan tenido en cuenta las aportaciones de las organizaciones de la sociedad civil. Además se lamenta que, mediante este código de conducta, las compañías tomarán el liderazgo en un asunto, el del control del discurso del odio online, que deberían ser liderado por las autoridades públicas, cuya actividad está sometida a un marco de responsabilidad democrática. Dejar el liderazgo de esta tarea a compañías privadas supone un claro riesgo para la libertad de expresión, en la medida en que los mecanismos de revisión de contenido de las compañías no están sometidos a un marco de responsabilidad claro y transparente³⁶. En el fondo, manifiestan los críticos, se trata de confiar en la responsabilidad de las empresas o, más bien, en que el interés de sus accionistas coincida, de alguna manera, con el interés social, y el respeto a la libertad de expresión, y los principios de igualdad, dignidad, etc. propios de los sistemas democráticos.

Otra cuestión que se plantea como problemática es que, en la medida en que el presunto discurso del odio sea eliminado por las compañías, y no se establezca un procedimiento de colaboración claro para el envío de dicho contenido a las autoridades públicas, los autores de las supuestas infracciones legales no serán sometidos a juicio ni castigo penal de ningún tipo³⁷.

De cualquier manera, no cabe duda, de que las instituciones europeas están seriamente preocupadas con el problema del discurso del odio online, y que están tratando de presionar a las compañías para que realicen una moderación más activa de las redes sociales.

35. http://europa.eu/rapid/press-release_SPEECH-16-2197_en.htm

36. <https://edri.org/edri-access-now-withdraw-eu-commission-forum-discussions/>

37. <https://edri.org/guide-code-conduct-hate-speech/>

4. LAS ESTRATEGIAS NO JURÍDICAS PARA COMBATIR EL DISCURSO DEL ODIO EN LAS REDES SOCIALES

En esta sección se presentarán, sin ánimo exhaustivo, las principales estrategias “no jurídicas” que se están llevando a cabo para responder al discurso del odio en general, y en las redes sociales en particular.

Buena parte de estas estrategias son directa o indirectamente financiadas por las instituciones públicas, y para algunos autores se plantean dos cuestiones importantes. Por un lado, si son realmente eficaces. Y por otro lado, la posibilidad de que se produzca una cierta vulneración de un supuesto principio de neutralidad del estado. Sin embargo, lo cierto es que incluso la mayoría de los autores que se oponen a las prohibiciones del discurso del odio consideran que las instituciones públicas no han de mantenerse neutrales ante este tipo de discurso extremo. Muy pocos autores defienden un principio de neutralidad estatal estricta; estos argumentan que una implementación equivocada o excesiva de ciertas estrategias puede producir efectos indeseados, como la marginación o la expulsión del debate público, de ciertos discursos, que pueden ser desagradables, pero legítimos. Para la gran mayoría, no cabe duda de que las instituciones públicas deben contribuir a la creación de un clima adecuado para un ejercicio de la libertad de expresión respetuoso con los principios de igualdad y no discriminación (Article 19 2015). Las administraciones pueden, a través de las declaraciones de los políticos y los funcionarios públicos, promover la empatía y el apoyo a los grupos víctimas del discurso del odio, difundiendo la idea de que este tipo de manifestaciones son dañinas para la convivencia. Este tipo de declaraciones son de vital importancia en situaciones de disputas entre comunidades étnicas en barrios o ciudades, o en contextos en los que las tensiones identitarias se recrudecen, como los periodos electorales. Otras iniciativas que las administraciones públicas pueden desarrollar para garantizar un clima adecuado para la igualdad y la no discriminación son la formación de los funcionarios públicos en estas cuestiones, la litigación estratégica, o la promoción del diálogo inter-religioso (Article 19 2015).

Las administraciones pueden, a través de las declaraciones de los políticos y los funcionarios públicos, promover la empatía y el apoyo a los grupos víctimas del discurso del odio, difundiendo la idea de que este tipo de manifestaciones son dañinas para la convivencia. Este tipo de declaraciones son de vital importancia en situaciones de disputas entre comunidades étnicas en barrios o ciudades, o en contextos en los que las tensiones identitarias se recrudecen, como los periodos electorales. Otras iniciativas que las administraciones públicas pueden desarrollar para garantizar un clima adecuado para la igualdad y la no discriminación son la formación de los funcionarios públicos en estas cuestiones, la litigación estratégica, o la promoción del diálogo inter-religioso (Article 19 2015).

En un plano más concreto, y en cuanto a iniciativas dirigidas específicamente contra el discurso del odio *online*, y en particular en redes sociales, se están llevando a cabo estrategias no jurídicas de diverso tipo.

- *estrategias de monitoreo y la investigación*, que pretenden conocer mejor la magnitud del problema del discurso del odio y explorar la posibilidad de utilizar sistemas de alerta temprana capaces de identificarlo automáticamente;
- *estrategias de presión a las empresas operadoras de las redes sociales*, encaminadas bien a que modifiquen sus políticas relativas al tipo de contenidos que se pueden compartir o que retiren contenidos específicos.

- *estrategias centradas en cambiar las percepciones y actitudes* de los usuarios de la Red frente al discurso del odio tanto en Internet en general como en las redes sociales, en particular. .

Entre este último tipo de estrategias, a las que este informe prestará más amplia atención por su mayor aplicabilidad en el nivel local, destacan dos grandes tipos: por un lado, las que consisten en *campañas*, que incluyen actividades de índole diversa y cuyo objetivo suele ser atacar los prejuicios y la intolerancia como raíz principal del problema del discurso del odio; y, por otro lado, las estrategias que se basan en *la formación y educación*, que incluyen la realización de talleres, seminarios y otros tipos de actividades destinadas a dotar a las personas, especialmente a los jóvenes, de las competencias necesarias para identificar y combatir el discurso del odio en internet.

4.1. Las estrategias de monitoreo e investigación

Este tipo de estrategias son importantes para entender el fenómeno del discurso del odio en las redes sociales de una manera integral: la extensión del problema, los diferentes tipos de odio, cómo y en qué situaciones se manifiesta con más virulencia, cuáles son los perfiles de los autores, qué tipo de lenguaje utilizan, qué plataformas son más propicias para la propagación del discurso extremo, qué efectos produce en otros usuarios, etc. Profundizar en el entendimiento del discurso del odio es valioso por el propio conocimiento, pero sobre todo, para aplicarlo a soluciones paliativas del problema, o al menos, a la prevención de sus consecuencias más dañinas.

Las tareas de monitoreo afrontan dificultades considerables, a pesar de los indudables progresos técnicos en la automatización de la identificación del discurso del odio. En primer lugar, por una cuestión conceptual. Como se señala en la primera parte de este informe, qué es y qué no es discurso del odio es objeto de controversia, y estas diferencias de concepto pueden generar resultados poco homologables. En segundo lugar, por la propia infinitud de las redes sociales. En tercer lugar, por los diversos tipos de contenido (fotografías, vídeos, audios, textos, etc.) a través de los cuales el discurso del odio puede manifestarse en las redes sociales. Y en último lugar, porque incluso cuando el medio es textual, existen multitud de combinaciones semánticas y códigos a través de las cuales canalizar el discurso del odio, sin necesidad de utilizar los insultos y expresiones que un programador puede prever (Ruiz et al. 2010). Una prueba fehaciente de la enorme dificultad de esta tarea son los errores de bulto cometidos por los algoritmos de filtrado automático de las propias operadoras de redes sociales, como Facebook, Twitter, Tumblr, etc, a pesar de que cuentan con los más potentes recursos técnicos y humanos.

Aunque apenas existen investigaciones sistemáticas sobre las conexiones entre el discurso del odio y el desencadenamiento de episodios de violencia, en el plano más práctico, las tareas de monitoreo pueden desempeñar una labor considerable como sistemas de alerta temprana en contextos inestables o en sociedades étnicamente divididas, para evitar sucesos como la trágica ola de violencia desencadenada tras las elecciones de Kenia de 2007 (Gagliardone et al 2015).

Por otra parte, en el campo de la investigación tienen especial interés los estudios y análisis, en el ámbito de la psicología social, sobre qué iniciativas son más efectivas contra el discurso intolerante. Actualmente son abundantes los programas dirigidos a combatir la intolerancia y los prejuicios, sin embargo muy pocos están basados en evidencias sólidas de su eficacia. Teniendo

en cuenta que buena parte de estas iniciativas son financiadas directa o indirectamente por las administraciones públicas, este campo de investigación adquiere una importancia capital para que esa inversión de dinero público sea socialmente rentable, o al menos, no contraproducente (Legault et al 2011).

Uno de los centros de referencia en el campo del monitoreo y la investigación, por la consistencia de sus estudios, es el *Centre for the Analysis of Social Media* del *think tank* británico Demos, que realiza frecuentes estudios no sólo sobre la extensión de diferentes formas de discurso del odio (islamofobia, sexismo, etc.), sino también sobre cómo funcionan las contranarrativas que se enfrentan al odio en las redes. En este campo de estudio, Facebook ha encargado a Demos la realización de una serie de investigaciones sobre cómo se producen y comparten los contenidos que combaten el discurso del odio, qué contenidos tienen más éxito y cuáles son potencialmente contraproducentes³⁸.

4.2. Las estrategias de presión a las redes sociales

Conscientes del rol crucial de las empresas operadoras de redes sociales como intermediarias de los mensajes extremistas e intolerantes, muchas organizaciones han optado por ejercer presión sobre las empresas para que actúen de manera más contundente contra el discurso del odio y para que sean más transparentes en relación con los resultados de sus mecanismos y políticas de moderación y retirada de contenidos. Este tipo de estrategias pueden enfocar directamente la presión hacia las redes sociales, o hacerlo indirectamente dirigiéndose a sus anunciantes.

Las estrategias de presión directa a las redes sociales se han llevado a cabo tanto desde organizaciones de la sociedad civil, como desde las instituciones públicas. Las organizaciones de la sociedad civil presionan a las compañías a través de campañas online, recogida de quejas, peticiones online, etc. Por su parte, algunos estados y organizaciones supranacionales presionan a las empresas para que cumplan con las leyes nacionales e internacionales y se comprometan en una moderación de contenido más activa, por temor a sanciones y exigencias de regulación más costosas (Article19 2015). Como ejemplo de resultado de la presión de las instituciones europeas, en mayo de 2016, Facebook, Twitter, Youtube y Microsoft firmaron un Código de Conducta con la Unión Europea, al que ya se ha hecho referencia en este informe.

Las campañas de presión a los anunciantes, por su parte, han logrado considerables éxitos. Un ejemplo de este tipo de iniciativas fue desarrollado conjuntamente en Gran Bretaña y Estados Unidos por WAM! (*Women, Action and the Media Group*) y *Everyday Sexism Project*, para exigir la retirada de contenido abusivo hacia las mujeres. Como consecuencia de una campaña de comunicación intensa, que incluyó la recogida de más 225.000 firmas a través de *change.org*, el uso del hashtag #FBrape o la publicación de un listado de compañías que se anunciaban en “páginas” sexistas en Facebook, lograron que 15 grandes empresas retiraran su publicidad de Facebook, y que la compañía se pusiera en contacto con las organizadoras de la campaña para buscar vías de cooperación, y se comprometiera públicamente a revisar sus condiciones de servicio y sus mecanismos de revisión de contenidos (Gagliardone et al. 2015).

³⁸<https://www.demos.co.uk/project/counter-speech-on-facebook-phase-2/>

4.3. Las estrategias orientadas a cambiar percepciones y actitudes

4.3.1. Las campañas contra el discurso del odio en internet

Atendiendo a su contenido y orientación específica, estas campañas pueden categorizarse en tres tipos: de sensibilización, afirmativas o restrictivas (Titley, Keen y Földi 2014). Antes de examinar esta tipología de campañas y algunos ejemplos específicos de cada uno de los tipos, merece prestar atención a la campaña *Viviendo juntos online: acción y campaña contra el discurso del odio* (campaña No Hate), lanzada por el Consejo de Europa (COE) en marzo de 2013, por ser una campaña integral, que incluye en sí misma una amplia variedad de estrategias contra el odio online. Esta campaña ocupa un lugar privilegiado en el ámbito de la lucha contra el discurso del odio en internet, ya que ha funcionado en los últimos tres años como un paraguas para una multitud de iniciativas en los niveles nacional, regional y local del entorno europeo.

a) La campaña No hate del Consejo de Europa

La campaña *No hate* se enmarca dentro del esfuerzo del COE de promover los derechos humanos en internet y está dirigida especialmente a los jóvenes europeos. La idea de realizar esta campaña surgió de los jóvenes representantes del Consejo Consultivo de la Juventud del COE después de la matanza de Utøya en Noruega, en el verano de 2012. Este ataque en el que perecieron 77 jóvenes puso de manifiesto los peligros asociados al discurso del odio en internet, que en algunos casos puede traducirse en violencia contra determinados grupos sociales. Sin embargo, más que abogar por la simple represión de este discurso, esta campaña adopta un marco de derechos humanos y, por tanto, reconoce también el derecho a la libre expresión, apostando por fomentar el desarrollo de la autorregulación por parte de los usuarios (Keen y Georgescu 2016).

La meta principal de esta campaña es movilizar a los jóvenes europeos, creando un movimiento social contra el discurso del odio en internet (*No Hate Speech Movement*). Para ello, esta campaña se marcó cuatro objetivos específicos: en primer lugar, reducir el nivel de aceptación del discurso del odio *offline* y *online*; en segundo lugar, prevenir y contrarrestar este problema con la educación en derechos humanos; en tercer lugar, sensibilizar sobre los riesgos que entraña la propagación del discurso del odio para la democracia y el bienestar de los jóvenes; y, en cuarto y último lugar, desarrollar y diseminar herramientas y mecanismos para denunciarlo, entre los que destacan el Observatorio de Discurso de Odio (*Hate speech Watch*), que permite que los usuarios remitan casos de discurso de odio que hayan identificado en las redes u en otros espacios digitales.

El alcance de esta campaña incluye internet en su conjunto, pero las redes sociales tienen una especial importancia, tanto como objeto de monitoreo de la campaña como como herramienta en su combate. En este sentido, en la campaña se han puesto en marcha actividades centradas en el ámbito específico de las redes sociales. Una de las iniciativas de la campaña que han tenido mayor repercusión han sido los *Action Days*, en los que los miembros de la misma actúan de manera intensificada durante un día para sensibilizar sobre un tema específico, por ejemplo el antisemitismo (9 de noviembre de 2016). En concreto, los *Action Days* prevén acciones específicas en los que los activistas vinculados a la campaña intervienen en las redes de cuatro maneras diferentes: en primer lugar, *expresando de solidaridad* con las personas víctimas del odio, a través de la difusión de la imagen del Action Day proporcionada por la campaña o añadiendo dicha imagen a la foto del perfil del usuario en las redes sociales; en segundo lugar, *denunciando*

el contenido de odio en el Observatorio de Discurso de Odio y añadiendo un contraargumento, que se comparte después en las redes; en tercer lugar, *compartiendo el contenido con contranarrativas* proporcionado por la campaña, en formatos como MEMEs, imágenes, infografías, etc.; y, en cuarto y último lugar, *organizando actividades offline*, con el objetivo de sensibilizar y educar al público joven sobre el problema objeto en el Action Day, siguiendo algunos recursos proporcionados en el marco de la campaña, y que también se dinamizan en las redes sociales.

Esta campaña se ha reproducido a nivel nacional, vehiculándose a través de los Comités Nacionales de Campaña. Estos Comités reúnen a las principales partes interesadas en cada país y siguen las Directrices proporcionadas por el Consejo de Europa para mantener la coherencia general de la campaña. Estos Comités están coordinados en su gran mayoría por los departamentos responsables de juventud a nivel de cada Estado, aunque en algunos casos están formados solo por ONGs y otras partes interesadas. Hasta el momento, se han formado más de 40 Comités, tanto en los países miembros del COE como en otros países como EE.UU., México o Marruecos. Los Comités han puesto en marcha diferentes campañas nacionales, que en muchos casos se han hecho eco de las iniciativas lanzadas desde el COE. En España, el Instituto de la Juventud (INJUVE) es el organismo encargado de la puesta en marcha y desarrollo de la misma, con la colaboración con la ONG Movimiento contra la Intolerancia³⁹. La presentación y el lanzamiento de la campaña se celebraron el 21 de marzo de 2014, con ocasión del Día Internacional para la Eliminación de la Discriminación Racial.

Esta campaña también se ha llevado al ámbito local, a través de la serie de actividades presenciales que han tenido lugar en diversas ciudades europeas. Además, la ciudad de Estrasburgo ha colaborado activamente con el COE, que tiene allí su sede, para poner en marcha un plan de acción local y un grupo de apoyo de la campaña. Entre otros eventos, este plan se materializó en la organización de actividades como el evento de danza *No Hate Sounds*, que tuvo lugar el 22 de mayo de 2016 en Estrasburgo⁴⁰. No obstante, hasta el momento no ha habido un movimiento específico de ciudades o red de ciudades que apoyase explícitamente la campaña.

b) Una tipología de campañas

Además de la campaña No Hate del COE, a lo largo del territorio europeo se han llevado a cabo otras campañas de alcance más restringido contra el discurso del odio en internet. Con respecto al contenido y orientación específica de estas campañas, estas pueden categorizarse, como ya se ha señalado al inicio de esta sección, en tres tipos: de sensibilización, afirmativas o restrictivas (Titley, Keen, y Földi 2014: 63).

En primer lugar, las **campañas de sensibilización** pretenden concienciar al público sobre el discurso del odio y sus consecuencias. En el ámbito europeo, un ejemplo es la campaña *betterinternetforkids.eu*, puesta en marcha por INSAFE (*European network of Awareness Centres*) e INHOPE (*Internet Association of Internet Hotlines*), dos redes activas en el ámbito de la seguridad en Internet y la persecución de las modalidades de cibercrimen que afectan a la infancia. Aunque más centrada en el ámbito de la seguridad en la Red, esta campaña financiada por la Comisión Europea también generó contenidos vinculados al discurso del odio, sobre todo a

39. Más información sobre la campaña en España:

<http://www.injuve.es/convivencia/noticia/campana-contra-la-intolerancia-en-internet>.

40. Más información en: <http://www.nohatespeechmovement.org/nohatesounds>

través de la celebración del “Día para la Seguridad en Internet” (*Safer Internet Day*)⁴¹. En la misma línea pero en el ámbito autonómico, destaca la campaña “Internet sin riesgos”, puesta en marcha por el Área de Juventud del Cabildo Insular de Tenerife y cuyo objetivo es sensibilizar contra la propagación del ciberodio y el ciberacoso machista⁴².

En segundo lugar, las **campañas afirmativas** tienen como objetivo presentar a las minorías de manera positiva al público general para prevenir comportamientos discriminadores. Estas campañas suelen tener en cuenta también el espacio *offline*, y se han puesto en marcha para cambiar la percepción de varios colectivos que suelen ser objeto de discurso del odio, como por ejemplo las personas LGTB, gitanas o musulmanas. En primer lugar, con respecto al colectivo LGTB destaca la campaña de ámbito global “Todos fuera” (*All out*), que pretende reivindicar los derechos de este colectivo a través de realización de micro campañas ante eventos como las Olimpiadas de Sochi o acciones en defensa de activistas LGTB como la campaña #FreeTheFive, que exigía la libertad para cinco activistas LGTB chinas⁴³. En segundo lugar, con respecto a las personas gitanas, otro ejemplo es la campaña internacional “¿Gitanos típicos?” (*Typical Roma?*), puesta en marcha en 2010 en Macedonia, Albania y Bulgaria, entre otros países, y cuyo objetivo era desmontar los prejuicios con respecto a las personas de este colectivo⁴⁴. En ámbito español, resalta también la campaña #YoNoSoyTrapacero, que protestaba contra la imagen negativa de los gitanos que difunde la Real Academia Española (RAE)⁴⁵. En tercer lugar, con respecto a las personas musulmanas, destaca la campaña británica “El Islam es Paz” (*Islam es peace*), centrada en proporcionar información sobre el Islam y las personas musulmanas, que desmonte los abundantes prejuicios sobre esta religión y sus practicantes⁴⁶.

Una variante interesante de las campañas afirmativas, son las basadas en contranarrativas. Las contranarrativas son mensajes que ofrecen una alternativa positiva a la propaganda extremista o que pretenden deconstruir o deslegitimar las narrativas extremistas (Silverman et al. 2016: 15). Desarrollar contranarrativas contra el discurso del odio en Internet es una de las prioridades de la campaña *No Hate* para el periodo 2016-2017. Asimismo, las campañas basadas en contranarrativas contra el odio se están potenciando desde la UE, a través de varios instrumentos de financiación⁴⁷. En efecto, la puesta en marcha de este tipo de campañas e iniciativas se incluye como uno de los objetivos del Código de Conducta en materia de incitación ilegal al odio en Internet, firmado entre la Comisión Europea y Facebook, Twitter, Youtube y Microsoft. En relación al potencial de esta herramienta, el británico *Institute for Strategic Dialogue* ha elaborado varios estudios y materiales de sensibilización sobre las contranarrativas. Uno de ellos es el informe “El impacto de las contranarrativas” (*The Impact of Counter-narratives*), que afirmaba después de un estudio piloto de un año de duración, que “la exposición a puntos de vista alternativos puede promover el pensamiento crítico o plantar una ‘semilla de la duda’ que más adelante madure en un cambio de alternativas o comportamientos” (Silverman et al., 2016: 44). De este instituto surgieron varias iniciativas financiadas por Facebook, como la “Iniciativa para el valor cívico en lí-

41. Más información en: <https://www.saferinternetday.org/web/sid/resources/gallery>

42. Más información en: <http://www.internetsinriesgos.com/wp-content/uploads/2013/11/folleto-ciberativista.pdf>

43. Más información en: <https://allout.org/es/campanas-destacadas/>

44. Varios vídeos de la campaña disponibles en: <https://www.youtube.com/user/typicalroma>

45. Esta y otras campañas de sensibilización.

46. Más información en: <http://www.islamispeace.org.uk/>

47. En el coloquio anual sobre derechos fundamentales organizado por la Comisión Europea en 2015, “Tolerance and respect: preventing and combating Antisemitic and anti-Muslim hatred in Europe”, se destacaba la importancia de promover contranarrativas emanadas de la sociedad civil. Más información sobre el Coloquio en http://ec.europa.eu/justice/events/colloquium-fundamental-rights-2015/index_en.htm.

nea” (*Online Civil Courage Initiative*), en la que se anima a compartir historias personales con el objetivo de combatir el extremismo y el discurso del odio en Internet⁴⁸, además de un Manual y una Caja de Herramientas sobre cómo elaborar contranarrativas⁴⁹.

En tercer lugar, las **campañas restrictivas** tienen como objetivo recoger información sobre páginas u acciones en línea basadas en contenido intolerante y tomar medidas para restringir dicha actividad. En este ámbito destaca la fundación International Network Against Cyberhate (IN-ACH), que está formada por organizaciones nacionales que recogen quejas y denuncias relativas a la discriminación en Internet, como por ejemplo el Movimiento contra la Intolerancia en España. Especialmente activos en el ámbito del antisemitismo, esta red coorganizó en abril de 2016 la primera “Conferencia Internacional sobre el antisemitismo en línea”, celebrada en Jerusalén⁵⁰. Asimismo, en 2016 publicó el informe “Devolverles al mar” (*Kick them back into the sea*) sobre el discurso del odio contra los refugiados⁵¹. Las organizaciones nacionales miembros de esta entidad también han puesto en marcha campañas que pretende aumentar las quejas y denuncias en este ámbito. Un ejemplo reciente es la campaña internacional “Get the Trolls Out”, puesta en marcha por varias organizaciones y apoyada por el COE cuyo objetivo es restringir el discurso del odio antisemita en Internet⁵². En el marco de esta campaña se generaron interesantes recursos, como la guía “How to counter hate speech on Twitter”⁵³, así como iniciativas que pretenden estigmatizar a los abusadores como el “Trol del mes”.

4.3.2. Las estrategias contra el discurso del odio en Internet basadas en la educación y la formación

Otro grupo de estrategias para combatir el odio en Internet son las basadas en la educación y la formación. Gran parte de estas estrategias entran dentro de la educación en derechos humanos, que incluye la educación para la diversidad y la interculturalidad, y están orientadas principalmente a los jóvenes.

Hasta el momento, se han puesto en marcha varias iniciativas educativas en los niveles europeo, nacional y local. Por un lado, en el marco de la campaña No Hate del COE destaca la elaboración de la publicación “Bookmarks”, un manual para combatir el discurso de odio online a través de la educación en derechos humanos. Este manual propone una serie de actividades educativas para contrarrestar este tipo de discurso (Keen y Georgescu : 2014: 21-133). Por otro lado, además de en el nivel europeo, se han puesto en marcha otras iniciativas contra el discurso del odio en Internet basadas en la educación y la formación en los niveles nacional y local, que en muchos casos siguen las recomendaciones del manual Bookmarks. Sin ánimo de exhaustividad, estas iniciativas pueden clasificarse en dos grandes grupos: por una parte, las que están orientadas a educar y formar a la población en general, para el desarrollo de lo que se ha conceptualizado como “ciudadanía digital”; y, por otra parte, las iniciativas destinadas a dotar a personas ya concienciadas y activistas con las aptitudes necesarias para actuar más eficazmente en el medio online.

48. Más información en: <https://www.facebook.com/onlinecivildcourage>

49. Más información en: <http://www.counternarratives.org/about-us>

50. Más información sobre la conferencia y sus recomendaciones en http://www.inach.net/detail.html?tx_news_pi1%5Bnews%5D=10&cHash=3bf353f3d766835f68a466d2f8c5b170.

51. Informe disponible en: http://inach.zone35.net/fileadmin/user_upload/Refugee_Report20161.pdf

52. Más información en <http://www.getthetrollsout.org/>

53. Disponible en: <http://stoppinghate.getthetrollsout.org/>

a) El pensamiento crítico y la reflexión como herramientas clave contra el discurso del odio en Internet

Desde la irrupción primero de Internet y después de las redes sociales en el mundo de la comunicación, las personas han pasado de ser solo receptores de mensajes a ser también emisores. Por tanto, se ha puesto de manifiesto que las estrategias educativas deben adaptarse a esta nueva realidad, promoviendo la recepción crítica de los mensajes, pero también el empoderamiento en la creación de contenidos (Hoechsmann y Poyntz 2012). Para lograr esta puesta al día, se ha propuesto la noción de “ciudadanía digital”, que pretende adaptar el concepto de educación para la ciudadanía, incorporando al mismo el conocimiento y aptitudes necesarias para interactuar en el medio digital (Gagliardone et al. 2015: 46). Estas aptitudes hacen referencia a la necesidad tanto de un pensamiento crítico que permita filtrar los contenidos que se reciben, como de una reflexión autocrítica antes de compartir contenidos propios o ajenos.

Algunas iniciativas contra el discurso del odio se han centrado en trabajar estas cuestiones, principalmente con jóvenes, ya que aunque todos los sectores de la sociedad generan, difunden y reciben contenidos con discurso del odio, estos son especialmente vulnerables por su uso intensivo de las redes. En efecto, en una presentación del director de investigación sobre redes sociales del *think tank* Demos en 2012 afirmaba que un elemento clave en la lucha contra el discurso del odio es dotar de herramientas a los jóvenes para desarrollar un pensamiento crítico adaptado a la era digital⁵⁴. En este sentido, cuestionar la información que aparece en Internet y aprender a buscar evidencias que apoyen las opiniones expresadas es central para proteger a los más jóvenes contra el “populismo digital” o formas diversas de discurso del odio.

En línea con el concepto de ciudadanía digital, reforzar el pensamiento crítico y la alfabetización mediática se están promoviendo desde diversos foros, en los ámbitos global y europeo. Por un lado, el programa de educación de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) tiene como una de sus áreas estratégicas de trabajo la educación para la ciudadanía mundial, que pretende, entre otras cosas, fomentar un “sentido de pertenencia a una humanidad común, compartiendo valores y responsabilidades, empatía, solidaridad y respeto de las diferencias y la diversidad” (Gagliardone et al. 2015). Entre los objetivos específicos de alcanzar una ciudadanía mundial, se incluye el desarrollo de capacidades y habilidades técnicas necesarias para usar las tecnologías digitales, así como el conocimiento y habilidades para buscar, analizar, evaluar e interpretar textos procedentes de medios de comunicación, crear mensajes en medios y reconocer su influencia social y política (Hoechsmann y Poyntz 2012).

Por otro lado, los ministros de educación de la UE adoptaron en una reunión informal en marzo de 2015 la denominada “Declaración de París sobre la promoción de la ciudadanía y los valores comunes de libertad, tolerancia y no discriminación a través de la educación”. En esta declaración se identificaban una serie de acciones en el ámbito educativo para promover la libertad de expresión, la inclusión social y el diálogo intercultural⁵⁵. Además, se marcaba como objetivo “reforzar el pensamiento crítico y la alfabetización mediática e informacional, en particular en el uso de Internet y las redes sociales, de cara a desarrollar resistencias a todos los tipos de discriminación y adoctrinamiento”⁵⁶.

54. Carl Miller (2012), “Research perspective”, discurso pronunciado en la conferencia “Tackling hate speech. Living together online” el 26 y 27 de noviembre de 2012 en Budapest. Disponible en http://hub.coe.int/c/document_library/get_file?uuid=8544adcf-f707-4bc1-8484-c6c55311cd1a&groupId=10227.

55. https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/images/1/14/Leaflet_Paris_Declaration.pdf

56. Ibid, p. 3

En lo que respecta a las iniciativas específicas basadas en mejorar estas competencias para prevenir y actuar contra el discurso del odio en Internet, se pueden mencionar varios ejemplos. En primer lugar, cabe destacar el proyecto “Prevención del discurso del odio en Internet” (*Prevention of online hate speech*) del Centro Europeo Wergeland, situado en Noruega y especializado en desarrollar programas de educación para la ciudadanía democrática. En el marco de este proyecto, se llevaron a cabo talleres y formaciones que daban a los participantes, educadores, estudiantes adolescentes y padres, herramientas para prevenir el discurso del odio y el extremismo. Estas formaciones se basaron en el manual Bookmarks del COE antes mencionado⁵⁷.

En segundo lugar, otra iniciativa interesante es el “Pack educativo contra la discriminación digital”, elaborado en el marco del proyecto ICUD⁵⁸. Este pack está destinado a públicos diversos (niños, jóvenes, padres, educadores, profesores, activistas, miembros de ONG’s, responsables políticos, académicos, entre otros) e incluye herramientas pedagógicas resultado de los talleres realizados con jóvenes y que está disponible en 6 idiomas, entre ellos castellano y catalán.

En tercer lugar, otra iniciativa que vale la pena mencionar es el proyecto pedagógico realizado en el marco de la campaña No Hate por la asociación cultural belga Loupiote “NO HATE - Ce qui vous regarde”⁵⁹. Aunque más centrado en el tema del acoso escolar a través de las redes, en el marco de este proyecto se realizaron una serie de talleres y formaciones con jóvenes y educadores de la región de Bruselas, que incluían el desarrollo de herramientas pedagógicas destinadas a combatir el ciberodio y el ciberacoso entre los jóvenes⁶⁰. La herramienta principal resultado del proyecto es un cortometraje desarrollado junto con un grupo de adolescentes que muestra las consecuencias del discurso del odio en las redes sociales, a través de la experiencia de un joven que sufre los comentarios homófobos y el acoso de sus compañeros de instituto. Además, este cortometraje está acompañado de varios vídeos de corta duración sobre temas vinculados como la libertad de expresión, el anonimato en Internet y la educación en los medios de comunicación, así como un vídeo musical sobre la propagación del odio en línea⁶¹.

b) Formando a ciberactivistas: algunas estrategias específicas dirigidas a la formación de colectivos ya concienciados

Además de las estrategias mencionadas en el apartado anterior, también se han llevado a cabo iniciativas destinadas específicamente a empoderar a las personas ya concienciadas o victimizadas con aptitudes específicas que les permitan actuar más eficazmente en línea. En particular, en los últimos años ha cobrado importancia la necesidad de empoderar a las propias víctimas del discurso del odio en Internet. De hecho, empoderar y apoyar a las víctimas del discurso y de crímenes de odio en internet es una de las prioridades de financiación del programa de Derechos, Igualdad y Ciudadanía de la UE para el periodo 2014-2020⁶².

57. <http://eng.theewc.org/Content/What-we-do/Completed-projects/Prevention-of-online-hate-speech>

58. Más información sobre el proyecto y acceso al pack en el enlace: <http://digitaldiscrimination.eu/pack/>

59. http://www.culture-enseignement.cfwb.be/index.php?eID=tx_nawsecured1&u=0&file=fileadmin/sites/cult_ens/upload/cult_ens_super_editor/cult_ens_editor/documents/News/Loupiote_NoHate.pdf&hash=ed1276c37b8dbfaa6864a9ee565d15fbc7a14b26

60. El video está disponible en francés en el siguiente enlace: <https://vimeo.com/111615933>

61. <https://loupioteasbl.wordpress.com/ce-qui-vous-regarde-no-hate/>

62. <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52016SC0158&from=EN>

Empoderar a activistas es el objetivo principal del proyecto *Gente joven combatiendo el discurso del odio en Internet*, que forma parte de la campaña No Hate y que pretende formar a jóvenes y asociaciones juveniles con las competencias necesarias para reconocer y actuar contra el discurso del odio e involucrarles en el lanzamiento de la campaña. El objetivo de las formaciones es compartir experiencias y estrategias exitosas, para aumentar el impacto de las actividades de blogueros y activistas. En el marco de este proyecto, se llevaron a cabo varios talleres y seminarios sobre diferentes temas, el último de ellos sobre cómo combatir el discurso del odio sexista⁶³.

Siguiendo la misma filosofía, y en el ámbito español, cabe destacar, por un lado, el taller “Cómo actuar ante el discurso del odio en Internet”, y el curso “Derechos Humanos y Discurso del Odio”, que incluía una unidad específica sobre estrategias para combatir el discurso del odio en Internet, realizados en el marco del Proyecto Online contra la Xenofobia y la Intolerancia en medios digitales (Proxi), puestos en marcha por el Institut de Drets Humans de Catalunya y United Explanations titulado, en los que se impartía formación sobre herramientas prácticas para identificar comentarios que vulneran los derechos humanos e intervenir eficazmente en los foros de los medios digitales⁶⁴. Por otro lado, en el marco del proyecto Redes Contra el Odio, se formó un grupo de cibervoluntariado compuesto por 15 jóvenes activistas, encargado de vigilar y denunciar la existencia de discurso de odio en Internet, así como de identificar situaciones de ciberacoso contra personas LGTB⁶⁵. En el ámbito local, destaca la mesa redonda “#Prouromors 2.0” sobre herramientas virtuales para desmontar estereotipos realizada en el marco de la Estrategia Antirumors de Barcelona y celebrada en junio de 2015 en esta ciudad, donde daban diferentes herramientas para luchar contra la rumorología “virtual”.

63. Más informació en: <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680667bcd>

64. http://www.observatorioproxi.org/images/pdfs/Programa_taller_Proxi_003_VF.pdf

65. <http://www.felgtb.com/redescontraodio/>

CONCLUSIÓN

Aunque sea complicado cuantificar de manera integral el problema, existe una sensación generalizada a nivel internacional de que el discurso del odio en las redes sociales está preocupantemente extendido. El odio, la ira y la agresividad se han convertido en moneda común en las redes sociales, provocando daño emocional a sus destinatarios y contribuyendo a la estigmatización y deshumanización de ciertos colectivos. Se le atribuye incluso la posibilidad de actuar como desencadenante de episodios de violencia en contextos conflictivos y sociedades étnicamente divididas. Las instituciones públicas nacionales e internacionales llevan años tratando de encontrar soluciones, o al menos remedios paliativos para un problema que parece descontrolado por las propias particularidades de Internet y las redes sociales: la sobreabundancia comunicativa, el anonimato, la transnacionalidad, etc.

Sin embargo, aunque existe cierto consenso sobre la extensión del problema, el concepto de discurso del odio sigue siendo discutido y es objeto de debate a nivel internacional. Es un concepto complejo, porque pone en juego principios fundamentales de los sistemas democráticos como la igualdad, la dignidad humana, y la libertad de expresión, cuya concepción no es idéntica en todos los contextos sociopolíticos. Por esta razón, es un concepto que puede ser implementado con objetivos más o menos legítimos.

El término ‘discurso del odio’ se utiliza para referirse a las expresiones que incitan a la discriminación o la violencia por motivos de odio racial, xenófobo, orientación sexual, u otras formas de intolerancia, pero también para referirse, en términos más amplios, a aquellas expresiones que fomentan la hostilidad a través de prejuicios o intolerancia. En los medios de comunicación, se ha generalizado el uso del concepto para referirse a una mezcla heterogénea de expresiones que incluye manifestaciones tan diversas, y que merecen enfoques muy distintos, como el acoso a famosos, las amenazas a individuos o colectivos por intolerancia, el discurso ofensivo contra el poder, la apología del terrorismo, o la blasfemia religiosa.

De cualquier manera, el objeto de este informe, sin ánimo de comprometerse con una u otra definición, son las expresiones que, en un sentido amplio, fomentan e incitan a la violencia, la discriminación o la hostilidad contra individuos o colectivos por motivos racistas, xenófobos, religiosos, de género, orientación sexual, discapacidad y otras formas de intolerancia.

Para sistematizar las diferentes respuestas e iniciativas que se llevan a cabo para combatir este fenómeno, se han clasificado en estrategias jurídicas y estrategias no jurídicas. Además, se ha prestado atención individualizada al rol de las empresas operadoras de redes sociales por su capacidad normativa y restrictiva como intermediarios en el ámbito privado.

Las estrategias jurídicas, que buscan la represión y sanción del discurso del odio, suponen un importante dilema por afectar a un derecho esencial para las democracias como la libertad de expresión. Los argumentos a favor y en contra de prohibir ciertas formas de discurso del odio pueden ser igualmente convincentes. Por ello, aunque la legislación penal emane del Estado, es importante tener en cuenta las implicaciones de este dilema como punto de partida para establecer una estrategia responsable frente al discurso del odio.

La última reforma del Código Penal en 2015 afectó a los delitos de discurso del odio, tratando de adaptar la legislación española a la normativa europea y de corregir las deficiencias de las que adolecía: falta de proporcionalidad, demasiado margen para la interpretación, y falta de adecuación al principio de intervención mínima del derecho penal. Como es habitual en asuntos de esta índole, no hay consenso entre los expertos sobre si los objetivos se han logrado.

Además de las dudas que plantea el recurso a las estrategias jurídicas por el conflicto que supone con el ejercicio de la libertad de expresión, esta vía se antoja insuficiente por dos razones. En primer lugar, porque la gravedad de las sanciones de prisión la convierte en una estrategia escasamente aplicable, más allá de algunos casos extremos. Y en segundo lugar, porque genera cierta incertidumbre por el escaso desarrollo jurisprudencial de la materia, y por la posibilidad de que tenga efectos contraproducentes.

Es innegable que el del discurso del odio un campo de actuación controvertido para las autoridades. Por este motivo, el rol de los operadores de redes sociales como intermediarios podría ser crucial. Así lo ha considerado la Unión Europea, que ha firmado un Código de Conducta con Facebook, Twitter, Youtube y Microsoft, en materia de incitación ilegal al odio en Internet, para lograr un compromiso de implicación en la moderación y la retirada de contenidos que atenten contra sus normas de uso y las legislaciones nacionales. Sin embargo, este tipo de iniciativas ha recibido importantes críticas por traspasar la responsabilidad de una cuestión sensible para la libertad de expresión a compañías privadas, y cuyos procedimientos de revisión y moderación no están sometidos a un marco de responsabilidad claro y transparente.

Existe un consenso considerable entre los expertos en que las estrategias “no jurídicas” son las que plantean menos objeciones. Quizás la asignatura pendiente de este tipo de estrategias, que en gran medida son directa o indirectamente financiadas por las instituciones públicas, sea la evaluación de su eficacia y rentabilidad social.

Se han desarrollado iniciativas y proyectos muy diversos. Se han llevado a cabo estrategias de monitoreo e investigación, no sólo como métodos de conocimiento, sino también como sistemas de alerta temprana en situaciones conflictivas. También se han desarrollado estrategias de presión a las empresas operadoras de las redes sociales, solicitando la modificación de sus políticas de autorregulación o exigiendo la retirada de contenidos específicos. Y sobre todo, se han puesto en marcha numerosos proyectos de intervención social con el objetivo de cambiar las percepciones y actitudes frente al discurso del odio.

Entre este último tipo de estrategias destacan las campañas, que incluyen actividades diversas con el objetivo de atacar los prejuicios y la intolerancia como raíz del problema del discurso del odio, y que pueden categorizarse en tres tipos: de sensibilización, afirmativas o restrictivas (Titley, Keen y Földi 2014). Mención especial merece la campaña *No Hate*, promovida por el Consejo de Europa, por la repercusión alcanzada y por ser una campaña integral, que incluye una amplia variedad de estrategias, entre las que destaca el desarrollo de contranarrativas. La contraargumentación a través de un discurso alternativo y deslegitimador del discurso del odio es una de las estrategias en las que se han depositado mayores expectativas por parte de las instituciones públicas y de las propias redes sociales.

Por último, y también con el objetivo de cambiar las percepciones y actitudes frente al odio, tienen especial importancia las estrategias educativas y formativas, ya sean dirigidas al de-

sarrollo de lo que se ha denominado como “ciudadanía digital”, o dirigidas a la formación de personas ya concienciadas, que puedan ejercer de activistas en el espacio online. Como es habitual en el campo de la intervención social, la educación y la formación generan grandes expectativas por su potencial de combinar de manera óptima la responsabilidad con la libertad y la participación. El reto, para no quedarse en el terreno de las buenas intenciones, está en cómo dotarlas de contenido concreto que logre los resultados deseados.

BIBLIOGRAFÍA

- Aguilar García, Miguel Ángel, ed. 2015. *Manual Práctico Para La Investigación Y Enjuiciamiento de Delitos de Odio y Discriminación*. Barcelona: Generalitat de Catalunya Centre d'Estudis Jurídics i Formació Especialitzada.
- Article 19. 2015. *Hate Speech' Explained A Toolkit*. Article 19. Londres. Disponible en https://www.article19.org/data/files/medialibrary/38231/Hate_speech_report-ID-files--final.pdf
- Bustos Gisbert, Rafael. 2015. "Libertad de expresión y discurso negacionista", en *Libertad de Expresión y discursos del odio*, Miguel Revenga Sánchez (dir.), *Cuadernos de la Cátedra de Democracia y Derechos Humanos* nº 12. Universidad de Alcalá.
- Castells, Manuel. 2012. *Redes de Indignación y Esperanza: Los Movimientos Sociales En La Era de Internet*. Alianza Editorial.
- Gagliardone, Iginio, Alisha Patel, y Matti Pohjonen. 2014. "Mapping and Analysing Hate Speech Online: Opportunities and Challenges for Ethiopia". University of Oxford and Addis Ababa University. <http://pcmlp.socleg.ox.ac.uk/sites/pcmlp.socleg.ox.ac.uk/files/Ethiopia%20hate%20speech.pdf>
- Gagliardone, Iginio, Danit Gal, Thiago Alves, y Gabriela Martinez. 2015. *Countering Online Hate Speech*. Programme in Comparative Media Law and Policy, University of Oxford.
- Gascón Cuenca, Andrés. 2015. "La Nueva Regulación Del Discurso Del Odio" a El Ordenamiento Jurídico Español: La Modificación Del Artículo 510 CP. *Cuadernos Electrónicos de Filosofía Del Derecho*, no. 32.
- Heinze, Eric. 2014. Nineteen arguments for hate speech bans- and against them. Disponible en: <http://freespeechdebate.com/en/discuss/nineteen-arguments-for-hate-speech-bans-and-against-them/> (Consultado el 23/11/2016)
- Hoechsmann, Michael., y Stuart R. Poyntz. 2012. *Media Literacies : A Critical Introduction*. Wiley-Blackwell.
- Jubany, Olga, y Malin Roiha. 2016. "Backgrounds, Experiences and Responses to Online Hate Speech: A Comparative Cross-Country Analysis"
- Jubany, Olga, Malin Roiha y Arlette Martínez. 2016. "Online Hate Speech in Spain. Field-work Report. PRISM. Preventing, Redressing & Inhibiting Hate Speech in New Media". Universitat de Barcelona. Fundamental Rights and Citizenship Programme of the European Union
- Kahn, Robert. 2015. Rethinking the Context of Hate Speech, en *First Amendment Law Review*; U of St. Thomas (Minnesota) *Legal Studies Research Paper* No. 15-20. https://works.bepress.com/robert_kahn/8/

Keen, Ellie, y Mara Georgescu. 2016. *Bookmarks - A Manual For Combating Hate Speech Online Through Human Rights Education*. Estrasburgo: Council of Europe. <http://site.ebrary.com/lib/uacj/docDetail.action?docID=10961376>.

Legault, Lisa, Jennifer N. Gutsell, y Michael Inzlicht. 2011. "Ironie Effects of Antiprejudice Messages: How Motivational Interventions Can Reduce (but Also Increase) Prejudice". *Psychological Science*, 22, 1472-1477. DOI: 10.1177/0956797611427918

Ministerio del Interior de España.

— 2015. Protocolo de actuación de las fuerzas y cuerpos de seguridad para los delitos de odio y conductas que vulneran las normas legales sobre discriminación. Disponible en: <http://gestionpolicia.diversidad.org/PDFdocumentos/PROTOCOLO%20ODIO.pdf>

— 2016. "Informe 2015 Sobre Incidentes Relacionados Con Los Delitos de Odio En España". Disponible en: <http://datos.gob.es/catalogo/informe-2015-sobre-incidentes-relacionados-con-delitos-de-odio-espana>

Moretón Toquero, María Aranzazu. 2012. "El «ciberodio», La Nueva Cara Del Mensaje de Odio: Entre La Cibercriminalidad Y La Libertad de Expresión." *Revista Jurídica de Castilla Y León*, no. 27.

Morozov, Evgeny. 2012. *El Desengaño de Internet: Los Mitos de La Libertad En La Red. Imago Mundi*. Barcelona: Ediciones Destino.

Phillipson, Gavin. 2015. "Hate Speech Laws: What they should and shouldn't try to do", *Revue générale du droit* (www.revuegeneraldudroit.eu), Etudes et reflexions 2015, número 13.

Revenga Sánchez, Miguel. 2015. "Los discursos del odio y la democracia adjetivada: tolerante, intransigente, ¿militante?", en *Libertad de Expresión y discursos del odio (op.cit)*.

Rey Martínez, Fernando. 2015. "Discurso del odio y racismo líquido", en *Libertad de Expresión y discursos del odio (op. cit)*.

Rodríguez Izquierdo, Myriam. 2015 "El discurso del odio a través de Internet", en *Libertad de Expresión y discursos del odio (op. cit)*.

Ruiz, Carlos, Pere Masip, Josep Lluís Micó, Javier Díaz-Noci, y David Domingo 2010 "Conversación 2.0 y democracia. Análisis de los comentarios de los lectores en la prensa digital catalana". *Comunicación y Sociedad*. Vol. XXIII, Núm. 2, 2010, pp.7-39.

Silverman, Tanya, Christopher J Stewart, Jonathan Birdwell, y Zahed Amanullah. 2016. "The Impact of Counter-Narratives. Insights from a Year-Long Cross-Platform Pilot Study of Counter-Narrative Curation, Targeting, Evaluation and Impact."

Teruel Lozano, Germán M. 2015. "La libertad de expresión frente a los delitos de negacionismo y de provocación al odio y a la violencia : Sombras Sin Luces En La Reforma Del Código Penal." *Indret: Revista Para El Análisis Del Derecho*, no. 4.

Titley, Gavan, Ellie Keen, and László Földi. 2014. "Starting Points for Combating Hate Speech online. Three Studies about Online Hate Speech and Ways to Address It."

Van Spanje, J. y C. de Vreese. 2015. "The Good, the Bad and the Voter: The Impact of Hate Speech Prosecution of a Politician on Electoral Support for His Party." *Party Politics* 21 (1). SAGE Publications: 115–30. doi:10.1177/1354068812472553.

Vázquez Alonso, Víctor J. 2015. "Libertad de expresión y religión en la cultura liberal: de la moralidad cristiana al miedo postsecular", en *Libertad de Expresión y discursos del odio (op. cit.)*.

Vives Antón, Tomás. 2015. "Sobre la apología del terrorismo como 'discurso' del odio", en *Libertad de Expresión y discursos del odio (op. cit.)*.