



*El Instituto Novact de Noviolencia es una asociación sin ánimo de lucro dedicada a la transformación de conflictos y la construcción de la paz. Durante más de 20 años, hemos trabajado en construir una paz positiva a través de la noviolencia en la región Euromediterránea. Nuestro compromiso es impulsar cambios en contextos frágiles y afectados por conflictos, creando políticas transformadoras que promuevan la paz, defiendan el espacio cívico, generen seguridad humana, justicia ecológica, justicia de género, justicia digital y prevengan abusos corporativos. Nuestra meta es asegurar que las decisiones que afectan al futuro de las comunidades sean tomadas por las propias comunidades. Novact quiere estar cerca de los movimientos sociales mediante una estrategia de protección integral, entendiendo que trabajar en la fuerza colectiva de las personas es crucial para la transformación social.*

Autoría: Irina Samy Curucull y Bernat Aragó Navarro

Coordinación: Núria Millán Iniesta

Con el apoyo de:



El contenido de este informe es responsabilidad exclusiva de Novact y no refleja necesariamente la opinión del Ministerio de Derechos Sociales y Agenda 2030.



Esta obra está sujeta a una licencia de reconocimiento-NoComercial 4.0 Internacional Creative Commons.

# Prólogo

Laia Serra Perelló,  
abogada experta en violencias digitales.

## La rendición de cuentas de las empresas tecnológicas:

La responsabilidad corporativa de las empresas hacia los derechos humanos alcanzó un hito con la aprobación de los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas en 2011. Su Principio 5 establece que las empresas, deberán guiarse por la diligencia debida para identificar, prevenir, mitigar y rendir cuentas. La aplicación de estos estándares ha sido desigual, siendo que las empresas tecnológicas hasta hace bien poco, han quedado fuera del escrutinio público. La proliferación de las violencias digitales contra las mujeres y otros colectivos, junto con la expansión de la Inteligencia Artificial han actualizado la urgente necesidad de su rendición de cuentas. En su último informe sobre moderación de contenidos online de 2023, la Agencia Europea para los Derechos Fundamentales (FRA), concluía que la misoginia era la forma más transversal de violencia digital detectada en las 4 plataformas analizadas.

A nivel europeo, no fue hasta mayo de 2016 que por primera vez la Unión Europea llamó al orden las empresas tecnológicas para atajar el desbordante fenómeno de los discursos de incitación al odio y a la discriminación. Mediante el denominado *Code of Conduct*, empresas como Facebook, Microsoft, Twitter y YouTube inicialmente y posteriormente Instagram, Snapchat, o TikTok asumieron una serie de compromisos de proactividad en el monitoreo de ese tipo de contenidos y de retirada dentro de las 24 horas de los contenidos lesivos. Una de las problemáticas de ese acuerdo, es que estando basado en la Decisión Marco 2008/913/JAI, estaba enfocado en los discursos racistas y xenófobos, desatendiendo los discursos lesivos para las mujeres.

Las instituciones internacionales y europeas no han priorizado el abordaje de ese tipo de discurso, prueba de ello es que, hasta la fecha no existe un *Code of Conduct* específico. Hasta el 2016 el Consejo de Europa no publicó su memorándum *Combating Sexist Hate Speech*, hasta el 2018 la Relatora de las Naciones Unidas sobre la violencia contra las mujeres no publicó su informe monográfico sobre estas, hasta el 2021 el GREVIO, comité que vela por la aplicación del Convenio de Estambul, no publicó su Recomendación General sobre la dimensión digital de las violencias, y hasta ese mismo año, no se propuso una Directiva Europea sobre violencias contra las mujeres que obligará a los Estados Miembro a regular como delito algunas infracciones digitales.

La falla sistemática de las empresas tecnológicas a la hora de prevenir las violencias hacia ellas en sus espacios online, sobre todo hacia ciertos sectores de mujeres especialmente vulneradas, no se explica por una cuestión de imposibilidad técnica, sino por una falta de voluntad y compromiso. La impunidad de estas violencias ha generado una movilización feminista cada vez más amplia, en la que las activistas, académicas y organizaciones que trabajan en los derechos tecnológicos y en el antirracismo, han sido punta de lanza. Sus iniciativas se han centrado en visibilizar la envergadura del fenómeno, su impacto en la calidad democrática y en los proyectos de vida de mujeres y niñas, la diversidad de violencias existentes, la ausencia de una respuesta institucional a la altura y la falta de rendición de

cuentas de las empresas tecnológicas, que en contra partida obtienen enormes beneficios a costa de nuestros datos.

Una iniciativa relevante en este sentido ha sido la de la *World Wide Web Foundation* que en 2021 organizó una serie de laboratorios para buscar soluciones contra las violencias de género digitales. Fruto del diagnóstico que acuñaron, en el Foro Generación Igualdad de las Naciones Unidas, empresas referentes como Facebook (Meta), Google y TikTok anunciaron la asunción de compromisos para erradicar esas violencias. Un año después, la *World Wide Web Foundation* junto con otras organizaciones como *Glitch*, concluyeron que el resultado había sido defraudador y llegaron a la conclusión colectiva de la necesidad de crear un mecanismo global de rendición de cuentas - *accountability* de las empresas tecnológicas.

Algunas de las notas clave que ese mecanismo de rendición de cuentas debía tener, era, el partir de la idea de que existe un vínculo indisoluble entre la igualdad de género y la calidad y fortaleza de las democracias, el hecho de que la diligencia debida aplicable a las violencias de género también lo era a las digitales; la centralidad de la experiencia de las víctimas; el tener en cuenta que el daño causado por esas violencias tiene una dimensión individual pero también colectiva o la necesidad de poner en el centro los sectores de mujeres especialmente dañados por esas violencias. El mecanismo se inspiraba en un criterio medioambiental universalmente asumido, coloquialmente denominado “quién contamina, paga”.

La articulación de la exigencia de un mecanismo global de rendición de cuentas de esas empresas coincidió en el tiempo con la tramitación del Reglamento europeo de Servicios Digitales (DSA), que entró en vigor en noviembre de 2022. Esa norma marca el fin de la etapa denominada de la auto- regulación por parte de las propias empresas y por primera vez marca un modelo de regulación con obligaciones y sanciones. Dentro de las obligaciones de las empresas, está la de la evaluación de los riesgos sistémicos, la de la transparencia y la de permitir el escrutinio institucional y público. Las obligaciones se estructuran en función de la envergadura de las empresas, en la línea de lo que planteaba la *World Wide Web Foundation*, es decir, que cuanto más lucro, más responsabilidades.

La Comisión Europea y los Estados Miembro enfrentan un reto de dimensiones democráticas, como es el de implementar esa norma. Está por ver su capacidad de poner coto a unos gigantes tecnológicos, que precisamente, basan gran parte de su modelo de negocio en la amplificación de contenidos, amplificación que se alimenta de la violencia y de la polémica. Según publicaba recientemente el *Center for Countering Digital Hate*, en su nueva etapa como X, la empresa habría rehabilitado 10 perfiles tóxicos, que estimaba que reportarían a la empresa 19 millones de dólares al año en publicidad. Por su parte los feminismos también enfrentamos el reto mayúsculo de formar parte de los procesos de fiscalización de esas empresas, para que nuestras necesidades colectivas sean tenidas en cuenta y para que en esta reñida contienda, no se pierdan derechos y libertades por el camino, en nombre de nuestra protección.

# Índice

---

<b>0</b>	<b>Introducción</b>	<b>6</b>
<b>1</b>	<b>Metodología</b>	<b>7</b>
<b>2</b>	<b>El discurso de odio</b>	<b>8</b>
	<i>¿Qué es el discurso de odio?</i>	8
	<i>Discurso ¿o delito? de odio</i>	10
	<i>Conflicto de intereses y derechos</i>	11
<b>3</b>	<b>El discurso de odio en las redes sociales</b>	<b>13</b>
	<i>¿Cómo se materializa el discurso de odio en Twitter?</i>	13
	<i>¿Quién se esconde detrás de estos ataques?</i>	15
	<i>La legitimización del discurso de odio</i>	16
<b>4</b>	<b>El caso de Twitter</b>	<b>19</b>
	<i>¿Un reflejo de la realidad?</i>	19
	<i>Discurso de odio: políticas internas y sanciones</i>	20
	<i>El odio y los algoritmos</i>	21
	<i>Crisis de moderación y materialización de los cambios</i>	21
<b>5</b>	<b>Combatir el discurso de odio en Twitter</b>	<b>24</b>
	<i>Actuar ante el discurso de odio</i>	24
	<i>Reacciones y respuestas</i>	24
	<i>Denuncias</i>	25
	<i>Seguridad online y traslado al plano físico</i>	28
	<i>Efectos del discurso de odio</i>	28
	<i>Redes de apoyo y cuidados</i>	30
<b>6</b>	<b>Recomendaciones</b>	<b>32</b>

## 0. Introducción

---

El discurso de odio<sup>1</sup> encuentra un lugar cada vez más amplio dentro de nuestras sociedades, en particular en el espacio digital, donde se multiplica de forma exponencial y con un carácter especialmente agresivo. Sin embargo, esta tendencia puede atribuirse a una dinámica propia del conjunto de la sociedad, que encuentra su razón de ser en factores como el auge de la derecha y la extrema derecha en el contexto europeo y que resulta en una polarización de los discursos.

A pesar de tratarse de una tendencia generalizada en las redes, Twitter, ahora llamado X, destaca como plataforma estrella en la que el discurso de odio prolifera particularmente, dado que la red nace con ánimo de convertirse en una plaza pública en la que existe interacción abierta entre personas con discursos, opiniones y experiencias distantes. La toma de control por Elon Musk ha contribuido a que Twitter, siga siendo uno de los espacios digitales populares con menos control.

Las personas atacadas por los discursos de odio suelen pertenecer a colectivos vulnerabilizados socialmente por razón de género, origen étnico, orientación sexual o religión, entre otros. En relación con el discurso de odio racista, según datos del Observatorio Español de Racismo y la Xenofobia (OBERAXE), los contenidos de discurso de odio en línea en España se concentran en las personas de origen magrebí, migrante, la infancia que viaja sola, personas de origen asiático y gitano.

Esta presencia prominente del discurso de odio en línea también ha generado el nacimiento de cada vez más iniciativas locales, estatales y europeas para analizar e intentar combatir este discurso de odio. Además, salvando las distancias de la dificultad de detección e identificación de culpables, la proliferación de estos discursos se encuentra cada vez más perseguida y criminalizada por los marcos legales nacionales y europeos.

No obstante, queda mucho camino por recorrer hasta conseguir espacios digitales seguros y libres de odio para todas las personas, y con este motivo se ha realizado la presente investigación. En ella, se pretende analizar el discurso de odio poniendo el foco en los perfiles atacantes, cediendo espacio para amplificar las voces de personas que han recibido odio en las redes y poniendo de relieve las estrategias de resistencia personales y comunitarias y generando recomendaciones dirigidas a la plataforma y a las instituciones para la prevención, detención, sanción y reparación ante los discursos de odio en línea.

---

1 Por finalidades prácticas y de facilitación de la lectura, la terminología usada a lo largo de este informe es “discurso de odio”. Sin embargo, la acepción terminológica más adecuada a esta práctica sería “discurso de incitación al odio”, debido a que este término pone el foco en la intencionalidad discriminatoria del mismo.

# 1. Metodología

---

Esta investigación se ha llevado a cabo principalmente mediante un análisis cualitativo de datos. Se han entrevistado a seis personas, todas ellas mujeres. Cinco de las entrevistas se han centrado en la vivencia personal en relación al uso de las redes sociales, Twitter en concreto, y han sido realizadas a personas activistas o personalidades públicas en diferentes ámbitos de los derechos sociales. La sexta entrevista fue realizada a una persona con perfil investigador. Son las siguientes:

**Sara Cuentas** es periodista, investigadora social y activista antirracista y feminista decolonial, forma parte de la Red de Migración, Género y Desarrollo.

**Sarahi Boleko** es la actual presidenta de SOS Racismo Madrid, militante y activista antirracista y feminista por la justicia social.

**Daniela Ortiz** es artista y activista decolonial y antirracista, en su trabajo analiza el poder colonial, capitalista y patriarcal.

**Miriam Hatibi** es consultora de comunicación, escritora y activista, trabaja por combatir la islamofobia y promover el feminismo interseccional.

**Najat Driouech** es filóloga y actualmente diputada al Parlamento de Cataluña y secretaria de Derechos Sociales, Ciudadanía y Vivienda en Esquerra Republicana de Cataluña.

**Luce Prignano** es investigadora y coordinadora del laboratorio tecnocientífico Heurística y experta en sistemas y redes complejos, con el objetivo de analizar los comportamientos sociales en Twitter y esclarecer las dinámicas del funcionamiento de los algoritmos de la red.

Las entrevistas han sido realizadas mediante ritmos consentidos y consensuados en todo momento con el fin de evitar la revictimización. Una retribución económica simbólica ha sido otorgada a aquellas personas entrevistadas que no son funcionarias públicas. Con esto, lo que se pretende es la puesta en valor no solo del tiempo, sino de los conocimientos de las personas entrevistadas dado que, a menudo, se extrae conocimiento de forma gratuita a las personas en condición de racialización o cuyos derechos han sido vulnerados, desvalorizando el tipo de conocimientos y sabiduría que poseen. La retribución económica es un reconocimiento y agradecimiento reivindicativo en el contexto actual y por ende necesaria.

Aunque en la actualidad la plataforma sobre la cual se basa esta investigación se llame “X” de manera oficial, por cuestiones prácticas y teniendo en cuenta la frecuencia con la que se dan cambios significativos en la red, en esta investigación se ha optado por el uso del antiguo nombre de la plataforma, “Twitter”, a excepción de cuando se quiere remarcar la diferencia entre el antes y el después del cambio de nombre y la evolución que este conllevó.

El análisis transversal de la investigación ha sido fruto de los inputs de las entrevistas y del conocimiento situado tanto de las personas entrevistadas como del equipo de investigación. Las fuentes externas consultadas están citadas a lo largo de la investigación.

## 2. El discurso de odio

---

### ¿Qué es el discurso de odio?

Poniendo como base el conocimiento situado de las personas entrevistadas, una definición posible sobre el discurso de odio es que se trata de una **violencia directa o indirecta, verbal o física que consiste en un señalamiento claro hacia una persona o grupo de personas por lo que representan en la sociedad**. Los ataques verbales fruto del discurso de odio tienen una tendencia agresiva y destructiva enfocada a atacar la persona, a la existencia y construcción de su ser y a su autoestima<sup>2</sup>. Sara Cuentas, una de las personas entrevistadas en la presente investigación, identifica uno de los elementos clave para definir los discursos de odio, y es que éstos tienen que ver con un relato que **“rompe con la legitimidad** de lo que significa esa persona a la cual estamos atacando”.

“El discurso de odio es cuando hay un señalamiento claro y agresivo hacia tu persona por el simple hecho no sólo de ser mujer, sino de ser mujer musulmana y que además lleva el *hijab*” -Najat Driouech.

En términos jurídicos, ciertamente **no existe una definición universal** de los discursos de odio expresada de manera clara y concisa en el marco del derecho internacional de los derechos humanos (DIDH). En este sentido, las definiciones existentes en los distintos códigos penales de cada país presentan importantes variaciones<sup>3</sup>.

En efecto, se trata de **un concepto que es objeto de debate** político, jurídico y académico, debido a las fricciones que genera tratar de establecer una frontera nítida entre el derecho a la no discriminación y a la igualdad por un lado y la libertad de expresión u opinión por el otro. También es cierto que **la palabra “odio” es un concepto subjetivo y emocional**, y muchas veces una expresión legítima, hecho que dificulta el entendimiento universal y a su vez puede ser relativamente manipulable<sup>4</sup>.

Sin embargo, aunque ningún tratado ni instrumento legal internacional ponga el foco en los discursos de odio, son muchos los organismos internacionales y regionales de derechos humanos que han profundizado en la cuestión y han demostrado una **base jurídica más que sólida** para darles cobertura. En primer lugar, en base al principio de igualdad y no discriminación, que rige cualquier interpretación posible de los instrumentos internacionales de derechos humanos y se ejemplifica en tratados regionales como el Convenio Europeo de Derechos Humanos en su artículo 14 sobre la prohibición de la discriminación<sup>5</sup>. Y, en segundo término, a través de los tratados específicos contra la discriminación en sectores concretos de la sociedad. Así es, la Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial (1965) declara punible en su artículo 4 las ideas basadas en “la

---

2 Definición formada a partir del conocimiento situado de las personas entrevistadas. Conocimiento situado es un concepto desarrollado por Donna Haraway que hace referencia a una postura epistemológica crítica en la cual se especifica desde qué punto de vista se parte, haciendo explícito el posicionamiento político de una manera ética, ya que los puntos de vista no pueden ser neutros.

3 Serra, L. (2018). Las violencias de género en línea. Pikara Magazine. p.10.

4 Cabo I., A. y García J., A. (2017). Contrólate en las redes! El discurso de odio en las redes sociales: un estado de la cuestión. Ajuntament de Barcelona: Área de Derechos de Ciudadanía, Cultura, Participación y Transparencia. Pp. 4-14.

5 Consejo de Europa. Convenio Europeo para la Protección de los Derechos Humanos y de las Libertades Fundamentales, 4 Noviembre 1950. Art. 14.

superioridad de una raza<sup>6</sup> o de un grupo de personas de un determinado color u origen étnico, o que pretendan justificar o promover el odio racial y la discriminación racial, cualquiera que sea su forma”<sup>7</sup>. Del mismo modo, la Convención sobre la Eliminación de todas las Formas de Discriminación contra la Mujer (1979) hace lo mismo con “toda distinción, exclusión o restricción basada en el sexo que tenga por objeto o resultado menoscabar o anular el reconocimiento, goce o ejercicio por la mujer”<sup>8</sup> en cualquier esfera.

En definitiva, existen sobradas interpretaciones autorizadas del derecho internacional de los derechos humanos que amparan la regulación de los discursos de odio en el marco legal internacional. Un claro ejemplo de ello lo hallamos en la Recomendación N°7 de la Comisión Europea contra el Racismo y la Intolerancia<sup>9</sup>.

A grandes rasgos, podríamos afirmar que **existen dos concepciones diferenciadas de los discursos de odio con aceptación a nivel internacional**. La primera, impulsada principalmente por el Consejo de Europa, que brinda una interpretación más amplia del concepto, englobando “**toda forma de expresión que difunda, incite, promueva o justifique el odio**”<sup>10</sup> basado en la intolerancia y los prejuicios discriminatorios<sup>11</sup>. Por otro lado, existe un segundo grupo de definiciones más restrictivas de los mismos, incluyendo únicamente aquellas expresiones o discursos que incitan de manera directa a la violencia o que se producen contextos de inestabilidad pudiendo desencadenar episodios violentos<sup>12</sup>. Cuando se usa el concepto en el presente informe se toma de referencia la primera definición, debido a su capacidad de englobar un abanico más amplio de formas de violencia discursiva que es necesario abordar, más allá de las cuestiones jurídicas.

Por otro lado, **las nuevas tecnologías y la eclosión de las redes sociales** han ocasionado, entre muchas otras cuestiones, el surgimiento de **nuevos escenarios donde los discursos de odio se agudizan y son reproducidos a gran escala**<sup>13</sup>. En consecuencia, los escenarios digitales han supuesto nuevos retos para hacer frente a los discursos del odio desde la perspectiva legal. Así pues, en los últimos 20 años se han desarrollado protocolos y observaciones a instrumentos de derechos humanos ya existentes con el fin de adaptar la legislación internacional a los nuevos contextos y ampliar su cobertura<sup>14</sup>.

6 Desde NOVACT se rechaza el uso del término “raza” debido a las connotaciones racistas que ha implicado su uso a lo largo de la historia y en la actualidad, implicando una jerarquización y supeditación entre distintos grupos de personas mediante la racialización –categorización– de las diferencias fenotípicas. Además, desde la perspectiva científica tampoco es correcto su uso pues no existe una diferencia genética significativa entre los seres humanos. En este párrafo, el concepto aparece citado de manera textual porque es así como figura en el documento original.

7 ONU: Asamblea General. Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial, 21 Diciembre 1965. Art. 4.

8 ONU: Asamblea General. Convención sobre la eliminación de todas las formas de discriminación contra la mujer, 18 Diciembre 1979. Art. 1-3.

9 Comisión Europea contra el Racismo y la Intolerancia (ECRI). Recomendación Núm. 7 de Política General de la ECRI sobre legislación nacional para combatir el racismo y la discriminación racial, del 13 de diciembre de 2002.

10 Consejo de Europa. Recomendación N° R (97) 20 del Consejo de Ministros a los estados miembro sobre el “discurso de odio”, 30 Octubre 1997. Pp. 107.

11 Consejo de Europa. Recomendación N° R (97) 20 del Consejo de Ministros a los estados miembro sobre el “discurso de odio”, 30 Octubre 1997. Pp. 107.

12 Cabo I., A. y García J., A. (2017). Contrólale en las redes! El discurso de odio en las redes sociales: un estado de la cuestión. Ajuntament de Barcelona: Área de Derechos de Ciudadanía, Cultura, Participación y Transparencia. Pp. 4-14.

13 ONU: Asamblea General. Consejo de Derechos Humanos. La discriminación racial y las tecnologías digitales emergentes: un análisis de los derechos humanos. Informe de la Relatora Especial sobre las formas contemporáneas de racismo, discriminación racial, xenofobia y formas conexas de intolerancia, 18 Junio 2020. Párr. 4.

14 A modo de ejemplo, el Comité para la Eliminación de la Discriminación Racial (CERD, por sus siglas en inglés), encargado de supervisar la aplicación de la Convención sobre la Eliminación de todas las Formas de Discriminación Racial, ha publicado desde 2002 varias Observaciones Generales que, entre otros aspectos, incluyen la perspectiva de los discursos de incitación al odio a través de Internet, tales como la Observación General N° 29, la N° 30, la N° 34 o la N° 35.

## Discurso ¿o delito? de odio

En el caso de las legislaciones española y europea, para que los discursos de odio sean considerados delitos es necesario que se den dos elementos. El primero, como es obvio, es que la motivación de tal acción esté basada en prejuicios con un claro objetivo discriminatorio o estigmatizante, pero el segundo, **requiere que la acción cometida esté tipificada como delito en el Código Penal**<sup>15</sup>. Sin embargo, los discursos de odio “legales” sí pueden actuar como pruebas de motivación de delito de odio en el proceso de enjuiciamiento de una infracción.

En cambio, si tal discurso incitase de manera directa a cometer actos de violencia o crímenes, en este caso sí sería tipificado como delito. Este es el caso, por ejemplo, de la reciente sentencia de la Audiencia Provincial de Málaga, que condena a siete personas de Melilla a dos años y medio de prisión por difundir, a través de las redes sociales, mensajes de odio hacia la infancia no acompañada con incitación explícita a la violencia, del estilo: “Si voy en coche atropello al que sea” o “tenemos que hacer limpieza en las calles con nuestros propios medios”<sup>16</sup>.

En definitiva, por un lado, existe un grupo de discursos de odio considerados punibles, y por lo tanto serían catalogados como delito por los párrafos 1 y 2 del artículo 510 del Código Penal. Son aquellos que promueven o incitan al odio o a la discriminación directa o indirectamente; la producción, elaboración, posesión o distribución de materiales orientados a tal fin; el enaltecimiento o justificación pública de un delito de odio; las acciones que suponen humillación, menosprecio o descrédito de una persona o grupo de personas en base a prejuicios discriminatorios; o la negación, trivialización o enaltecimiento de delitos de genocidio o de lesa humanidad. Por otro lado, otros discursos de odio de menor gravedad y sin una intencionalidad concreta de incitación al odio que quedarían amparados por la libertad de expresión y, por lo tanto, no tendrían consecuencias jurídicas. No obstante, que este grupo de discursos no sea considerado punible ni tipificado como delito por el Código Penal, no significa que no merezca ser combatido, ya sea por la vía legal o desde otras perspectivas, puesto que supone igualmente una grave amenaza para la convivencia y una contribución a la discriminación<sup>17</sup>. Sin embargo, la interpretación de la legislación en materia de discursos de odio no está exenta de debate y aún hoy en día es objeto de controversias a medida que se van creando precedentes judiciales.

Además, la última reforma del Código Penal –llevada a cabo en 2015 con el fin de amoldarse a las directrices europeas– **estipula como un agravante el hecho de que un delito de odio se lleve a cabo en plataformas digitales, medios de comunicación o a través de internet**, considerándose que el alcance puede ser mucho mayor<sup>18</sup>. Un claro ejemplo de ello fueron las oleadas de discursos islamófobos que se produjeron en España a través de las redes sociales tras los atentados de Barcelona y Cambrils en agosto de 2017, que convirtieron varias mezquitas del país en un blanco de ataques racistas<sup>19</sup>.

El Ministerio del Interior del Gobierno de España señala en un informe que **los delitos de**

15 Gobierno de España: Ministerio de Empleo y de Seguridad Social (2017). Legislación sobre los delitos de odio. Guía práctica. Subdirección General de Información Administrativa y Publicaciones. Pp. 19-42.

16 Vargas Martín, J. (25/09/2023). “Condenan a prisión a siete personas por incitar a la violencia contra menores migrantes en Melilla a través de Facebook”. Público.

17 Bazzaco, E.; García Juanatey, A.; Lejardi, J.; Palacios A.; Tarragona, L. (2017). “¿Es odio? Manual práctico para reconocer y actuar frente a discursos y delitos de odio”. Institut de Drets Humans de Catalunya; SOS Racisme.

18 Ley Orgánica 1/2015, de 30 de marzo, por la que se modifica la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal. Boletín Oficial del Estado, 77, de 30 de marzo de 2015. [https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2015-3439](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2015-3439)

19 Aragón, A. (20/08/2017). “Las mezquitas, blanco de ataques islamófobos tras el atentado de Barcelona”. Cadena Ser. [https://cadenaser.com/ser/2017/08/19/sociedad/1503156666\\_447678.html](https://cadenaser.com/ser/2017/08/19/sociedad/1503156666_447678.html)

**odio han crecido cerca de un 30%** desde 2020, mientras que destaca también un aumento de más de un 62% en otros incidentes relacionados con el odio que han sido puestos en conocimiento del sistema judicial<sup>20</sup>. Este mismo informe también esclarece dos cuestiones cruciales. La primera es que **casi la mitad de los delitos de odio que se producen en el Estado español se dan por motivaciones racistas o xenófobas**. La segunda es que los delitos racistas y xenófobos son los que más se dan en internet y en las redes sociales.

Los discursos del odio, independientemente de si incitan de manera directa a la violencia o no, tienen unos **efectos devastadores** tanto en las personas como en las comunidades. Difundir mensajes de odio, generalizaciones falseadas y estereotipos basados en prejuicios afecta de manera obvia a las personas a las que van dirigidos, tanto en el plano emocional como en la vulneración de derechos humanos, tales como la integridad, la identidad, la libertad religiosa o la igualdad y la no discriminación. Pero aún más allá, **estos mensajes son la antesala de los horrores** con los que el odio ha marcado la historia de la humanidad: Armenia, Auschwitz, Bosnia, Rwanda, Gaza y otros.

El caso de Myanmar es uno de los más paradigmáticos, ya que la difusión de los discursos de odio hacia la comunidad musulmana rohinyá a través de Facebook durante meses contribuyó a la violencia en el mundo real en la campaña de limpieza étnica por parte de las fuerzas de seguridad.

Lejos de ser inofensivo, el discurso de odio promueve un señalamiento y linchamiento público contra colectivos o comunidades determinadas. Como hemos visto, esta exposición puede acabar transformándose en desigualdades materiales, vulneración de derechos o episodios de violencia.

### *Conflicto de intereses y derechos*

Como se ha comentado anteriormente, la causa principal de que los discursos de odio estén en constante debate surge de la **colisión entre varios de los derechos humanos más básicos**: la libertad de expresión y el derecho a la dignidad humana, así como también el derecho a la igualdad y a la no discriminación. Esta cuestión ha sido objeto de muchos debates en el seno de las Naciones Unidas, pero sus conclusiones siempre han apuntado a la compatibilidad entre ambos preceptos<sup>21</sup>. En este sentido, como establece el artículo 19 del Pacto Internacional de los Derechos Civiles y Políticos (PIDCP), **la libertad de expresión no es un derecho absoluto**, sino que cuenta con las limitaciones de los otros derechos. El dónde se ubica la frontera es el punto de tensión entre las dos principales corrientes. La primera visión, de carácter más bien liberal, concibe el debate público como un mercado de ideas donde el Estado no debe interferir, mientras que otras visiones, de origen comunitarista, son más intransigentes con los discursos de odio y abogan por su restricción<sup>22</sup>.

Resulta evidente que los discursos de odio atentan contra valores como la inclusión o la diversidad y que representan un impedimento a la consecución de los derechos humanos. Tal como se ha señalado previamente, los efectos del odio en las sociedades pueden ser terribles

20 Gobierno de España: Ministerio del Interior (2022). Informe sobre la evolución de los delitos de odio en España 2022. Pág. 2-12.

21 Ejemplos de ello son la Recomendación General N° 35 sobre la lucha contra el discurso de odio racista o la Recomendación General N° 15 relativa al artículo 4 de la Convención, ambas del Comité para la Eliminación de la Discriminación Racial.

22 Marciani B., B. (2013). "El lenguaje sexista y el hate speech: un pretexto para discutir sobre los límites de la libertad de expresión y de la tolerancia liberal". Revista Derecho del Estado, (30), 157-198.

y su proliferación debe ser percibida como **una señal de alarma** que debería poner en alerta a todas las instituciones y organizaciones.

Sin embargo, para poder implementar políticas públicas destinadas a combatir los discursos de odio, **es necesario conocer las dinámicas del fenómeno y los medios por donde se propaga**, lo que requiere de un control, estudio y análisis en profundidad. En este sentido, y aunque en algunos casos las políticas deben orientarse a la penalización de según qué mensajes, la judicialización de los discursos de odio puede convertirse en un arma de doble filo y terminar por socavar la libertad de expresión a través de restricciones a la crítica política. De este modo, a partir de una concepción errónea de los discursos de odio, se han llegado a aplicar sanciones hacia personas o colectivos por publicar sátira política, por reivindicar sus derechos o por reprender y hacer frente a auténticos discursos de odio, entre otros. Algunos ejemplos de ello que ayudan a ilustrar lo mencionado serían, respectivamente, el caso de Cassandra Vera, condenada por publicar en Twitter chistes sobre Carrero Blanco<sup>23</sup>; el enjuiciamiento y/o prohibición de movimientos políticos como el movimiento palestino BDS, confundiendo el antisemitismo con el antisionismo<sup>24</sup>; o la identificación de la extrema derecha o las personas de ideología fascista como víctimas de los discursos de odio<sup>25</sup>.

23 Pinheiro, M. (01/03/2018). "Estos son los tuits sobre Carrero Blanco por los que la Audiencia Nacional condenó a prisión a Cassandra Vera". EIDiario.es.

24 El movimiento del BDS (Boicot, Desinversiones y Sanciones a Israel) es una campaña política surgida de la sociedad civil Palestina que rechaza la corriente política del sionismo y propone aislar política, cultural y económicamente al Estado de Israel con el fin de ejercer una presión para conseguir determinados objetivos políticos y el respeto de los derechos humanos del pueblo palestino. Varios activistas han sido juzgados en Europa, e incluso en algunos países se ha llegado a ilegalizar el propio movimiento, acusándolo de antisemita. Un ejemplo de ello en España fue el juicio contra varios activistas del movimiento BDS en València por llevar a cabo una campaña de boicot al concierto del cantante israelí Matisyahu en el marco del Festival Rototom. Fuente: Serra, H. (07/01/2019). "Piden cuatro años de cárcel por instar a un festival a que no contratase a Matisyahu". Público. <https://www.publico.es/sociedad/rototom-piden-cuatro-anos-carcel-instar-festival-no-contratase-matisyahu.html#analytics-noticia:contenido-enlace>

25 Un ejemplo de ello es la inclusión por parte de la Fiscalía General del Estado en su circular 7/2019 la "ideología nazi" como categoría protegida susceptible a ser víctima de los discursos de incitación al odio. Fuente: Castilla, K. (06/07/2019). "¿Nazis víctimas de delitos de odio?" El Periódico de Catalunya. <https://www.elperiodico.com/es/opinion/20190706/nazis-victimas-de-delitos-de-odio-articulo-opinion-karlos-castilla-7538596>

### 3. El discurso de odio en las redes sociales

---

#### *¿Cómo se materializa el discurso de odio en Twitter?*

El discurso de odio que se emite por las redes sociales consiste en una **denigración de la persona agredida, en el que, mediante un proceso de deshumanización, el agresor tiene por objetivo dañar la autoestima y silenciar la voz** de la persona. Suelen consistir en insultos y amenazas verbales y gráficas de un alto nivel de agresividad, que se pueden dar de forma masiva, tanto en mensajes públicos como de manera privada. Al tratarse de agresiones con motivación racista, más allá de la denigración, se habla de un “**cuestionamiento**” de la persona, ya que uno de los efectos del racismo es la puesta en duda de la propia existencia de la persona racializada, sus experiencias, su pertenencia, su conocimiento, etc. y lo que ésta representa en la sociedad.

Las causas identificadas por las mujeres entrevistadas son varias. En concreto, refieren haber sido atacadas por **señalar y denunciar el racismo, por religión, por defender causas sociales, por opinar sobre procesos políticos o por su condición familiar** como, por ejemplo, ser madre soltera. A pesar de la diversidad entre las personas entrevistadas, se identifica un nexo común en los ataques de odio, y es que estos **apuntan a una inventada “alteridad” de la persona que victimizan**. Esta alteridad se fundamenta en condiciones sociales personales que, organizadas por ejes de dominación social, interseccionan entre sí creando situaciones particulares de opresión. Algunos de los ejes que generan opresión y sirven para *alterizar* a la persona atacada son **el género, el racismo, el origen, la religión, la apariencia, el clasismo o el colorismo**<sup>26</sup>.

## Hacemos frente al discurso del odio

*“Aprovechan tu vulnerabilidad para tratar de quitarte del camino, porque el antirracismo es algo incómodo”*

**Daniela Ortiz**



---

<sup>26</sup> El término “interseccionalidad” fue acuñado por la abogada negra estadounidense Kimberlé Crenshaw en 1989 y hace referencia a la interacción entre diferentes factores sociales que definen socialmente a una persona, creando una situación de opresión o privilegio según su condición.

En las palabras de Sarahi Boleko, las personas atacantes hablan desde la supuesta pertenencia a un espacio y señalan la supuesta no-pertenencia de la persona a la que atacan, a menudo con inclinación racista y de género. De forma simplificada, se traduciría en un “yo pertenezco, tú no, y por esto mereces ser señalada y te señalo”. En esta exposición, se menosprecia la persona a la que atacan y su razón de ser, a partir de factores como su origen, opiniones, conocimientos o logros.

Es probable que, **cuantos más ejes de opresión interseccionen en una persona, más ataques de odio reciba, con más impunidad, y de mayor gravedad**. Esto se explica por el racismo, entendido como una estructura social y jerárquica que consiste en la opresión, dominación e inferiorización de un “grupo humano” por encima de otros. Estos “grupos humanos” se justifican mediante la categorización de las diferencias étnicas y fenotípicas entre personas y la jerarquización de estas diferencias. Es decir, la racialización de las diferencias crea grupos de personas “racializadas”: estas están separadas de las personas del grupo dominante, que es el blanco europeo y occidental.

Asimismo, cabe destacar que tanto la cantidad de mensajes de odio como la gravedad de los mismos se verán **condicionadas por el nivel de exposición pública** de una persona, convirtiéndose este en un factor más a tener en cuenta a la hora de hacer un análisis del discurso de odio en las redes sociales.

“Es un odio generalizado pero que se intensifica cuando cruza el género. La agresión se vuelve mucho más severa cuando es a mujeres a quien se interpela.” - Sarahi Boleko

Un factor clave de ataques de odio es pues, la **mera existencia de la persona y lo que su cuerpo, por ser negro, indígena, de mujer, o por llevar hijab, significa**. Esto quiere decir que, en muchos casos, la temática u opinión por la cual la persona es atacada no es el motivo principal de la agresión, ya que se tratan de ataques racistas y de género.

“Toda publicación que yo hago recibe ataques claramente machistas e islamófobos”, declara Najat Driouech. En el caso de los **ataques islamófobos** se observa un odio generalizado hacia el rechazo de la religión en sí y todo lo que evoca a ella. Según explica, el discurso de odio en formato de ataques islamófobos se extiende no solo a lo que publica Najat, sino también a cualquier persona que comparta o apoye una publicación suya, independientemente de si es musulmana o no.

“Lo que yo comparto en Twitter se recibe de otra manera si viene de un hombre blanco. Seguiría habiendo insultos, pero estos insultos no irían ligados a su dignidad personal e individual” - Sarahi Boleko

También son objeto de odio común aquellas **publicaciones que tratan temas que incomodan**. Algunos ejemplos macro son el racismo, el colonialismo o el patriarcado. En términos generales, estos debates o discusiones que señalan los privilegios en las desigualdades tienden a incomodar especialmente a personas blancas o hombres cis que no están familiarizados con los conceptos, y se sienten atacados al escuchar estos discursos de carácter crítico. Esta reacción hace referencia a la **“fragilidad blanca”** o a la **“masculinidad frágil”**, que consisten en el posicionamiento a la defensiva o de negación generado por la incomodidad de hablar sobre injusticia y desigualdad racial o de género.

Sin embargo, nuevamente, la persona que transmite estos mensajes incómodos y cómo es leído su cuerpo en la sociedad –si está racializado, es decir, si no es blanco–, es un factor clave en el cómo estos mensajes son recibidos. La **legitimidad discursiva** que se le otorga a un hombre blanco, por ejemplo, no es la misma que se le otorga a una mujer negra: generalmente se suele prestar más atención y dar **más crédito y reconocimiento al discurso de un hombre que de una mujer, de la misma manera que acontece con una persona blanca que con una persona en condición de racialización.**

### ¿Quién hay detrás de estos ataques?

Al tratar el discurso de odio en redes, pocas veces se ha puesto el foco en las personas atacantes, pues resulta difícil esbozar un perfil de las personas usuarias cuyas cuentas divulgan mensajes de odio sin entrar en generalizaciones, y más teniendo en cuenta el factor del anonimato en Internet.

Expertas y medios de comunicación han analizado en muchas ocasiones el comportamiento de ciertos “perfiles falsos” o de las cuentas automatizadas o *bots* y coinciden en que ni los perfiles falsos ni los *bots* son maliciosos de por sí. Por un lado, el anonimato en las redes sociales puede tener muchas razones inocuas: por ejemplo, personas que quieren separar un perfil público de carácter profesional de otro de uso personal para aficiones, cuentas temáticas especializadas... Por el otro, los *bots* también pueden resultar ser herramientas útiles con fines no discriminatorios, como cuentas compilatorias de información específica.

No obstante, es innegable que estos elementos pueden formar parte de campañas u operaciones de información con la finalidad de manipular la opinión pública, a menudo mediante el uso de discurso de odio. Según la experiencia de las personas entrevistadas, los ataques son frecuentemente emitidos desde **cuentas anonimadas**: perfiles sin nombre, apellidos o foto. Sarahi Boleko se refiere a estas cuentas secundarias como “cuentas alternativas”, es decir, cuentas que en su experiencia son usadas para manifestar odio libremente y sin sufrir consecuencias por ello. Denuncia que, Twitter, al tratarse de una empresa privada con ánimo de lucro, **no cuente con un proceso de verificación de perfiles obligatorio**, lo cual permite la creación de un número ilimitado de perfiles que pueden tener identidad anónima.

“Si recibo insultos de un usuario que se llama “user002003”, por ejemplo, es imposible para mí hacer un seguimiento de la cuenta, porque sé que va a desaparecer en unos meses, ya que los únicos tweets que comparte son insultos a colectivos muy determinados y con objetivos concretos.” - Sarahi Boleko

Por otro lado, a la tendencia de los discursos de odio en Twitter, causada hasta ahora por perfiles falsos o cuentas anónimas o pseudónimos, hay que añadirle una reciente **normalización de tales mensajes desde cuentas personales identificadas con nombres y apellidos**. Míriam Hatibi ha sido consciente de este cambio y explica que, mientras cinco años atrás recibía ataques casi exclusivamente de perfiles anonimados, en la actualidad se encuentra con ataques de perfiles vinculados a personas identificables, que emiten el discurso de odio “de forma súper naturalizada”. Esto nos indica un cierto **desacomplejamiento de ideologías autoritarias y de extremismos violentos** con un fuerte carácter anclado a la ultraderecha, que se materializa en discursos y expresiones racistas, islamóforas, machistas o LGTBfóbicas. Cabe decir que esta variación no es exclusiva de Twitter, o ni siquiera de las redes sociales o Internet, si no que puede atribuirse a una dinámica propia del conjunto de la

sociedad.

En cuanto a perfiles identificables que emiten odio, también se han reconocido personas provenientes de **contextos políticos considerados de izquierdas**. Esta experiencia la comparten Daniela Ortiz y Sara Cuentas, y aluden a personas conocidas de espacios sociales y de activismo.

“Yo siempre espero que estos discursos vengan del sistema, del contexto más opresor, pero no del propio entorno, dónde tú considerabas que estabas segura.” - Sara Cuentas

A pesar de tratarse en su mayoría de ataques independientes entre sí, las personas entrevistadas coinciden en que existe un **componente de organización** en cuanto a su forma de operar, y que usan las redes para **atacar de forma sistemática a personas de colectivos que suelen estar en situación de vulnerabilización**. En los comentarios, se expresa una posición polémica, a veces desde la corrección política, normalmente basada en estereotipos y radicalmente extrema. Los comentarios polémicos actúan como **polo de atracción para un alud de decenas de comentarios** de otros usuarios con un perfil más bajo (con menos popularidad e interacción), generando así un efecto puente que se retroalimenta y contribuye a la polarización de los discursos. Así, estos “perfiles puente” son cuentas que se ubican en márgenes ideológicos extremos pero que se presentan de manera más neutral o con apariencia moderada y cuidadosa en su expresión. Éstos se transforman en los intermediarios atrayendo otros perfiles que propagan discursos de odio de manera más directa o descarada. Lo hacen a través de comentarios en publicaciones de perfiles que se encuentran en el centro del espectro político, introduciendo posturas polémicas y racistas de manera más sutil.

### *La legitimización del discurso de odio*

La naturalización del discurso de odio y su materialización en los ataques de odio se encuentran amparadas por una realidad de impunidad. Algunos de los factores que favorecen esta impunidad, y sus directos responsables, son...

#### **Es responsabilidad de Twitter...**

- La posibilidad de crear perfiles anonimizados
- La posibilidad de crear un número ilimitado de cuentas sin verificar
- La posibilidad de abrir nuevos perfiles en caso de que la cuenta sea cerrada por denuncias
- La moderación de contenidos y la eliminación de los discursos de odio

#### **Es responsabilidad de la comunidad internacional...**

- Un marco jurídico digital que no regulariza empresas privadas como Twitter
- La falta de legislación ante los delitos y discursos de odio o de que la legislación sea jurídicamente vinculante

**Es responsabilidad social...**

- El crecimiento disparado de la presencialidad en la virtualidad, es decir, el aumento de la cantidad de tiempo invertido online.
- La falta de conciencia sobre cómo habitan los espacios virtuales.

**La dificultad o la falta de voluntad de las plataformas en el rastreo de perfiles** emisores de discurso de odio tiene un impacto directo sobre la forma de los mensajes de odio. Debido a estos factores, los mensajes toman un **carácter mucho más agresivo y virulento**.

“Por la distancia digital hay personas que se atreven a decirte cosas que jamás te dirían en persona”. -Daniela Ortiz

Además, el tipo de perfiles y colectivos **más susceptibles de ser atacados son aquellos que se ven más afectados por las desigualdades estructurales y por el racismo sistémico**: según Daniela Ortiz, los agresores saben que los ataques a este tipo de perfiles serán tratados con aún más impunidad, y que generalmente reciben menos apoyo o tienen menos acceso a recursos.

“Saben que un hombre blanco español con buenas condiciones económicas tendría acceso fácil a un abogado, no dudaría en denunciar y consecuentemente les caería un proceso judicial si lo atacaran.”- Daniela Ortiz

La impunidad ante el discurso de odio trae consigo la **legitimización del discurso de odio**. Una forma muy sencilla de legitimización es mediante la **cantidad de interacciones que una publicación recibe**, ya sea en forma de *likes*, retuits, citaciones o comentarios explícitamente a favor. “Si yo te ataco y mi publicación tiene 10.000 *likes*, yo tengo mayor legitimidad que tú, si solo tienes veinte *likes*. La legitimidad se da a través de las redes sociales y tiene que ver con esa cantidad numérica de visualizaciones y *likes*”, explica Sara.

## Hacemos frente al discurso del odio

*“La impunidad empieza en las redes. Después, esta misma gente va desacomplejando su discurso fuera de las redes.”*

**Miriam Hatibi**



En relación con la impunidad, Míriam Hatibi apunta que, el hecho de que haya más impunidad en las redes sociales no significa que el discurso de odio se quede en las redes sociales. Al contrario, **esta impunidad permite la desacomplejización del odio fuera de las redes sociales, es decir, aumenta el riesgo de que este odio se convierta en violencia física**: “Hay gente que hace este tipo de discurso de odio en las redes y a la vez se están organizando fuera de las redes para hacer ataques racistas, o forma parte de movilizaciones de extrema derecha. Nunca sabes si la persona te está insultando o basta, o si además resulta que está dentro de un grupo neofascista”, comenta.

La legitimización del discurso de odio conlleva la **deslegitimación no solo de discursos, sino también de la persona atacada**. Por definición, tanto la legitimización como la deslegitimación tienen un carácter público y socialmente consensuado o aceptado. Esto se vincula con las llamadas “**cancelaciones**”, que hacen referencia a la práctica de retirar el apoyo a alguien públicamente y especialmente en las redes sociales<sup>27</sup>. Es una tendencia que se suele realizar en masa como forma de expresar desaprobación y de ejercer presión social. Daniela Ortiz los llama “castigos públicos” y aunque en muchos contextos las cancelaciones constituyen una práctica necesaria, ella reflexiona sobre el uso irresponsable que se hace de los castigos públicos en contra de activistas: “Por redes se genera toda una narrativa que luego salta a los medios de comunicación. Una vez estás deslegitimada socialmente, es muy difícil defenderte y recibir apoyo”.

La **inmediatez con la circulan los mensajes** en Internet y en especial en Twitter, evade la reflexión individual ante una tendencia de deslegitimación y cancelación, dificultando la posibilidad de escuchar voces alternativas o de que la persona atacada tenga la oportunidad de defenderse públicamente.

“Es sumamente preocupante el peso que le damos a la virtualidad”. - Sara Cuentas

<sup>27</sup> Definiciones extraídas del diccionario Merriam-Webster.

## 4. El caso de Twitter

---

### *¿Un reflejo de la realidad?*

La plataforma de *microbloggeo* Twitter fue creada por Jack Dorsey en 2006 con la vocación de convertirse en una plaza pública, poder opinar libremente y conocer las opiniones de otras personas, pudiendo llevar a cabo debates e interactuar activamente con desconocidos. Con la evolución de la red, ésta ha ido incorporando como usuarios a personajes públicos, importantes figuras políticas, periodistas profesionales, medios de comunicación, asociaciones y todo tipo de perfiles institucionales. Actualmente quien no tenga un perfil en Twitter no forma parte de la vida pública: se ha convertido en el escaparate informativo y de opinión a escala mundial.

Hay que decir también que la descripción precedente es parcial, puesto que no existe una sola realidad en Twitter, sino que conviven varios sistemas paralelos en función del uso que desempeñan los usuarios. De hecho, según la propia compañía, los temas principales de interés serían las criptomonedas y la pornografía, representando esta última un 13% del total de los contenidos<sup>28</sup>.

Así pues, elementos como la publicidad, las redes de contacto que se generan en función de los intereses comunes entre grupos de usuarios y el número de seguidores desigual entre unos y otros perfiles generan distorsiones sobre la percepción de la vida real en Twitter, sin embargo, no deja de ser un espejo de la realidad. Ahora bien, se trata de un reflejo deformado, polarizado y no representativo.

### **El atractivo de Twitter**

La relación de algunas de las activistas entrevistadas más activas en Twitter con la plataforma podría definirse en una relación amor-odio. Mientras identifican puntos negativos como el límite de extensión de las publicaciones -la red admitía hasta inicios de 2023 publicaciones con un límite de 280 caracteres- y la dificultad de tratar temas complejos con pocos caracteres, las entrevistadas coinciden en la característica estrella de la red: la **posibilidad de tener interacción abierta con distintas comunidades y personas**. Para las activistas, esto es algo que tiene un valor especial al tratar temas políticos.

“La interacción es abierta y se da no solo con tus amigos y contactos. Entrar en contacto con gente que no conoces es interesante cuando haces trabajo político, sobre todo hacer difusión y plantear cuestionamiento. Pero por otro lado también puedes entrar en contacto con grupos de extrema derecha, por ejemplo”, explica Daniela.

---

28 Zilber, A. (26/10/2022). Porn, crypto more popular on Twitter as users lose interest in fashion, celebs: report. New York Post.

## Discurso de odio: políticas internas y sanciones

Ante la problemática que supone dibujar unos límites concretos en el momento de establecer unas políticas internas, la red social Twitter ha optado por instaurar una normativa de uso prohibitiva y la previsión de varios paquetes de sanciones para aquellos usuarios que la incumplan. Por lo menos en la teoría.

En efecto, el Centro de Ayuda de Twitter afirma que la red no va a permitir ataques o amenazas “por motivo de su raza<sup>29</sup>, origen étnico, origen nacional, pertenencia a una casta, orientación sexual, género, identidad de género, afiliación religiosa, edad, discapacidad o enfermedad grave”, que resumen como “categorías protegidas”. La misma empresa afirma comprometerse con garantizar el derecho a la libertad de expresión al mismo tiempo que especifica, uno a uno, algunos de los tipos de contenidos que vulnerarían las normas de uso. Entre ellos, se mencionan el uso de referencias a formas de violencia con intención de acoso, la incitación o glorificación a la violencia, los insultos o tópicos con fines de degradación o las expresiones que fomentan la deshumanización.

Para combatir semejantes mensajes, la política de Twitter anuncia **tres niveles de sanción** según el grado de infracción, tales como la reducción de la visibilidad del contenido en cuestión (eliminando el post, limitando su visibilidad y posicionamiento, restringiendo las interacciones del mismo, o impidiendo los anuncios), la solicitud al usuario para suprimir el contenido en casos leves, o en casos más graves, directamente la suspensión de la cuenta del emisor.

Sin embargo, de la teoría a la práctica hay un larguísimo trecho, y los mecanismos de Twitter para poner en marcha las medidas anteriores dejan mucho que desear, al menos en la actualidad. Hoy en día, **Twitter es una red con un control y moderación escasos**<sup>30</sup>, puesto a que la moderación de contenidos requiere de una denuncia previa, a diferencia de otras redes como Tik Tok o Instagram, que ejercen el control sobre los contenidos de manera más explícita y frecuente.

Teniendo en cuenta que la red se autodefine como “Twitter es lo que está pasando”, es normal que una de las características fundamentales de la plataforma y a la que las personas entrevistadas refieren sea **la inmediatez con la que se transmiten los mensajes**. Esta característica tiene aspectos positivos y negativos. Centrándonos en los negativos, información tergiversada o falsa puede difundirse a toda velocidad, dificultando las respuestas pausadas y críticas. Esto impulsa al posicionamiento inmediato y frecuentemente acrítico de centenares de personas ante cualquier cuestión y genera miedo al juicio en caso de no posicionarse o de posicionarse a contracorriente a la tendencia.

Junto con la creación de tendencias de masa, la inmediatez hace que Twitter se convierta en el **espacio perfecto para deslegitimar a una persona**. En las palabras de Daniela, “Te queman. Te queman y después esta confianza cuesta muchísimo de recuperar. Las relaciones de amistad política que has tenido se rompen porque se ha generado una narrativa y ya no se quieren acercar a ti porque estás quemada y temen que al acercarse se les queme a ellos también”.

29 Citación textual. Desde NOVACT se rechaza el uso del término "raza" como queda detallado en la nota al pie número 6 en la página 9.

30 AFP (18/04/2024). “Elon Musk aborda nueva política de moderación en Twitter durante un evento ante anunciantes”. France 24. <https://www.france24.com/es/minuto-a-minuto/20230109-miles-de-cuentas-restablecidas-en-twitter-amenazan-multiplicar-la-desinformacion-seg-un-experto>

## El odio y los algoritmos

Los mensajes discriminatorios y la crispación en la red no son un fenómeno nuevo en Twitter. Por el contrario, son muchas las razones que los expertos en comunicación han puesto encima de la mesa para tratar de explicar por qué esta plataforma promociona los comportamientos beligerantes u hostiles: la limitación a los 280 caracteres por publicación, la excesiva de búsqueda de la inmediatez, el efecto llamada de los mensajes y comportamientos agresivos, un efecto espiral hacia la transgresión perpetua (que implica que llamar la atención en la red requiera cada vez más una mayor transgresión) o la cuestión del anonimato<sup>31</sup>.

Aun teniendo en cuenta lo anterior, resulta erróneo pensar que Twitter está diseñado para difundir mensajes de odio y propagar la ira entre sus usuarios. Como cualquier otra plataforma social privada, Twitter está concebida para generar el máximo beneficio posible, que obtiene a través del grado de interacción de los internautas en una cuenta, también conocido como *engagement*. Cuanto más tiempo se destine a la red, más alto será el beneficio económico.

En este sentido, el entramado de fórmulas matemáticas que forman los llamados algoritmos tiene por objetivo maximizar el tiempo que cada usuario esté conectado a la red social, seleccionando de manera nada casual qué contenidos vemos, en qué orden y cuántas veces aparecen en nuestro timeline. Así, los mensajes que generan polémica, los discursos de odio, los contenidos que causan sorpresa a la audiencia, como el *clickbait*<sup>32</sup>, o que indignan a los usuarios tienden a contar con muchas más visualizaciones e impactos en la red que aquellos que generan indiferencia. Ya sea por estar de acuerdo con ellos, o por indignación, los mensajes de odio actúan como cebo para atrapar a las personas en la red y sigan consumiendo los contenidos de Twitter.

Es decir, no es que el algoritmo de Twitter promueva los discursos de odio, sino que tales discursos encajan con todos los requisitos para convertirse en contenidos virales y generar rédito. Sin embargo, debido a la opacidad de los engranajes que hacen funcionar tal algoritmo, existe siempre cierta duda o sospecha sobre los parámetros que definen la recomendación de contenidos a los usuarios dentro de la propia red y sobre por qué muchos de ellos incluyen mensajes discriminatorios.

Es importante destacar que, además, existen actores con intereses muy específicos y una clara voluntad política de aprovechar esta sintonía entre los algoritmos y los mensajes discriminatorios para explotar el máximo rendimiento de la red y multiplicar exponencialmente su impacto. Esto sitúa la moderación de la red en un rol clave para limitar y erradicar este tipo de discursos, o en su ausencia, corre el riesgo de convertir la plataforma en un altavoz del odio.

## Crisis de moderación y materialización de los cambios

Desde que Elon Musk tomó las riendas de Twitter, rebautizado ahora como X, se ha podido observar varios cambios de rumbo en la gobernanza de la plataforma, especialmente en lo que concierne la moderación. Mientras que Jack Dorsey, cofundador de Twitter y predecesor de Musk en la dirección ejecutiva de la empresa, sostenía ciertos esfuerzos para cuidar la

31 Bigas Formatjé, N. (13/03/2020). "¿Por qué Twitter es la red del odio?". Universitat Oberta de Catalunya (UOC). <https://www.uoc.edu/portal/es/news/actualitat/2020/165-twitter-red-odio.html>

32 El clickbait es una técnica de redacción que pretende llamar la atención y captar el máximo número de clics posibles a través de encabezados y descripciones sensacionalistas, y muy a menudo escondiendo partes clave de la información para generar interés.

reputación y la respetabilidad de la red social a través de un intervencionismo moderado en los contenidos y los usuarios, Musk ha seguido la dirección contraria. Precisamente Dorsey, quien había mantenido un punto de equilibrio entre la maximización de los beneficios y el cumplimiento de las políticas internas en relación con los discursos de odio, ha criticado duramente la gestión de Musk en varias ocasiones<sup>33</sup>.

El almacenamiento de una cantidad tan ingente de contenidos como es Twitter, junto a la perseverancia en la moderación de la red, acarrear unos altos costes económicos, que Musk se ha propuesto revertir. Tras el despido de 6.000 empleados de Twitter Inc. –que representaban cerca del 80% de la plantilla–, la moderación de contenidos ha sido la principal damnificada<sup>34</sup>. También los portales de transparencia y el libre acceso a la API (interfaz de programación de aplicaciones) han sido restringidos, lo que imposibilita estudiar a ciencia cierta los datos y los comportamientos de los usuarios de esta red social.

El deterioro de la moderación de contenidos se ha traducido en una **menor fiscalización de los discursos de odio**. Es decir, cada vez es menos frecuente que las cuentas que vulneran las políticas de Twitter sean sancionadas o bloqueadas, y menos son los impedimentos para que los contenidos hirientes, estereotipados o discriminatorios proliferen en X. Sin embargo, Musk ha defendido el nuevo rumbo de la moderación en la red como una política que permite la completa libertad de expresión, aunque no la libertad de alcance<sup>35</sup>. En otras palabras, los discursos de odio no serán sancionados o eliminados de por sí, aunque su alcance va a ser limitado.

En la práctica, el aumento de ataques violentos es uno de los cambios más significativos que se han podido observar desde el traspaso a Elon Musk. Esto también se vincula con el cambio de paradigma en la **verificación de cuentas mediante pagos**: otro de los cambios destacados es que cualquier usuario de Twitter puede obtener una verificación para su cuenta –el *tick* al lado del nombre de usuario, en la actualidad azul para personas y el que más credibilidad tiene socialmente, oro para empresas y gris para cuentas oficiales–, y esto hace que sus publicaciones tengan más visibilidad y sus interacciones sean premiadas económicamente. En palabras de Sarahi Boleko, “lo que hace Elon Musk en Twitter ahora es premiar la polémica”.

Esto se traduce en un aumento en la visibilidad de contenido violento y que, a menudo, actúa contra los derechos de colectivos concretos sin represalia alguna. Mientras que las personas entrevistadas coinciden en que el nivel de agresividad de los mensajes ha tomado una dimensión exponencial, Sarahi observa la bajada considerable de interacciones y audiencia a la que llegan desde su colectivo antirracista por haber tomado la decisión de no pagar para la verificación de su cuenta y la frustración que conlleva esta nueva limitación.

**“Está afectando el hecho de que personas con cuentas verificadas emitan discurso de odio, porque tienen más visibilidad, interpelan de forma directa con acusaciones falsas y fake news hacia colectivos”. - Sarahi Boleko**

33 Jiménez de Luis, Á. (01/05/2023). “El fundador de Twitter critica la gestión de Musk: “fue un error venderle la red social”. El Mundo. <https://www.elmundo.es/tecnologia/2023/05/01/644ef7fbc4d4d8f3378b4591.html>

34 Planas Bou, C. (12/04/2023). “Elon Musk ha despedido el 80% de la plantilla de Twitter”. El Periódico. <https://www.elperiodico.com/es/tecnologia/20230412/elon-musk-admite-comprar-twitter-85901012#:~:text=Musk%20calcula%20que%20Twitter%20vale.una%20fiesta%22%2C%20ha%20apuntado.>

35 Musk, E. [@elonmusk] (18/11/2022). “New Twitter policy is freedom of speech, but not freedom of reach. Negative/hate tweets will be max deboosted & demonetized, so no ads or other revenue to Twitter. You won't find the tweet unless you specifically seek it out, which is no different from rest of Internet.” Twitter. [https://twitter.com/elonmusk/status/1593673339826212864?s=20&t=AjYIOHeAsbo\\_OuP-tkU3Q](https://twitter.com/elonmusk/status/1593673339826212864?s=20&t=AjYIOHeAsbo_OuP-tkU3Q)

En definitiva, la falta de transparencia y claridad en los criterios que determinan qué contenidos se borran, cuáles no, y qué características reúnen las publicaciones que son promocionadas o se viralizan generan una desconfianza justificada hacia la moderación aleatoria de la red.

## 5. Combatir el discurso de odio en Twitter

---

### *Actuar ante el discurso de odio*

#### Reacciones y respuestas

Las formas de actuar frente a ataques de odio son muchas. Las personas entrevistadas compartieron sus saberes al respecto, admitiendo ante nada que la forma de gestionar los ataques cambia con el tiempo por distintos factores: el compartir con compañeras que han vivido situaciones similares, los años de experiencia usando Twitter y la situación vital en la que la persona se encuentre. No existe una forma única o universal, solo el compendio de experiencias, conocimientos y consejos aprendidos y compartidos, que pretenden desgranarse en los párrafos que siguen.

Cuando los mensajes de odio llegan en masa, hay quien opta por ignorar y eliminarlos, como hizo Sara siguiendo el consejo de compañeras y hay quien intenta responder a cada uno de los mensajes, como estuvo haciendo Daniela durante un tiempo, hasta que una compañera abogada le aconsejó que dejara de hacerlo. Me dijo: “nunca respondas a los mensajes, especialmente si están obsesionados con atacarte con cosas personales, porque esperan que tú les des explicaciones”, recuerda Daniela. Al **responder estos mensajes, es muy probable que se acabe brindando información adicional a la persona agresora**, cuya identidad es desconocida, lo cual puede poner en riesgo la seguridad de la persona atacada. Además, el **desgaste emocional y el consumo de tiempo** que supone leer y responder mensajes de odio no es de subestimar, en especial al tratarse de ataques en masa, frecuentemente anónimos y de una violencia exponencial.

Como herramienta recurrente, existe el **bloqueo de las cuentas** que emiten los ataques. A nivel práctico, bloquear a una persona significa que su perfil deja de existir para ti y tu perfil deja de existir para ella, aunque te busque a través de la plataforma. A nivel práctico, si un usuario bloquea a otro, significa que sus perfiles dejarán de ser visibles entre ellos de manera respectiva, impidiendo también la interacción e inhabilitando la cuenta en el motor de búsqueda de la aplicación. Asimismo, algunas de las otras opciones que brinda Twitter son **denunciar perfiles y denunciar comentarios** específicos, prácticas que, junto con el bloqueo, todas las personas entrevistadas han llevado a cabo. Sarahi, en concreto, explica que es algo que ha hecho de manera muy frecuente y que algunos de los comentarios que ha denunciado han sido eliminados por la plataforma.

Las personas entrevistadas también han compartido medidas “preventivas”, en concreto dos, que se suelen aplicar a partir de la experiencia de recibir mensajes de odio. Una es la **privatización del perfil**. Esta consiste en hacer que tu perfil de Twitter no sea público, y que por lo tanto solo las personas seguidoras pueden ver el contenido compartido. Y la otra es la **restricción de quién puede responder los tweets**. Es decir, limitar la posibilidad de comentar sobre el contenido que se publica a solamente aquellas personas que se etiqueten en la publicación. Esta función está habilitada desde julio de 2020<sup>36</sup> y es útil dado que gran parte de los ataques de odio se realizan mediante los comentarios. Esta es una herramienta que

---

36 Merino, M. (11/08/2020). “La opción de limitar quién puede responder a nuestros tuits llega por fin a todos los usuarios.” Genbeta. <https://www.genbeta.com/redes-sociales-y-comunidades/opcion-limitar-quien-puede-responder-a-nuestros-tuits-llega-fin-a-todos-usuarios>

Miriam Hatibi empezó a usar recientemente, ya que, además, explica que en los comentarios de una publicación se suelen iniciar conversaciones paralelas que en muchas ocasiones se retroalimentan en el discurso de odio.

“Ahora que se pueden limitar los comentarios y quien comenta, lo limito. Pero no hay una forma de prevenir que la gente deje de participar; siempre se pueden crear cuentas nuevas. Esto es lo malo.” - Sarahi Boleko

Sarahi apunta a la trampa de que, aunque se utilice esta última configuración, cualquier persona sigue pudiendo citar un tweet. La función de citar tweets consiste en escribir un nuevo tweet mencionando el tweet ya existente de una persona, que se ve trasladado como hipervínculo con imagen en el nuevo tweet. A modo práctico, **esta función reactiva la posibilidad de iniciar conversaciones en público y de acceso abierto**, y anula la supuesta intención de restringir comentarios que generen conversaciones de acceso abierto.

Mientras algunas voces apostan por la necesidad de que la plataforma digital verifique la identidad de las cuentas con documentación oficial, tal como han hecho redes como Facebook en algunos casos, pretender acabar con el anonimato en las redes podría resultar contraproducente. Por un lado, perjudicaría a muchos usuarios que no participan de la difusión de los discursos de odio. Por otro, la propia existencia de las redes perdería en cierto modo su razón de ser, ya que en muchos casos es precisamente el anonimato lo que muchos usuarios buscan en las plataformas en línea. Y por último y no menos importante, podría convertirse en una herramienta implacable en contextos de represión política y supondría un grave riesgo para activistas y defensoras de derechos humanos.

## Denuncias

Tras una denuncia mediante la plataforma, existe un proceso de seguimiento que Sarahi califica como “simple” y “poco efectivo”, ya que la misma persona denunciada puede abrirse nuevas cuentas en caso de que la inicial sea cerrada por Twitter.

“Si le cierran una cuenta, le cierran la siguiente, se vuelve a abrir otra y se la pueden cerrar a los seis meses, pero esta persona ya ha estado twitteando, y si no ha sido denunciada o si el contenido denunciado no ha sido valorado como negativo, puede seguir campando a sus anchas”, explica Sarahi.

Las denuncias en la plataforma se llevan a cabo a través de un formulario dentro de la propia aplicación donde se puede especificar el motivo de la denuncia: odio, abusos y acoso, discurso violento, protección de personas menores de edad, privacidad o contenidos inapropiados, entre otros. No obstante, la respuesta de la moderación de la red a tales denuncias es muy limitada y no van más allá de las sanciones previstas por Twitter comentadas anteriormente. De hecho, al verificar y comprobar los contenidos, la empresa puede calificar el contenido como inapropiado, pero considerar que no se han vulnerado los términos y condiciones de la red. En estos casos la moderación de la plataforma suele inhibir el mensaje del perfil del denunciante en vez de suprimirlo definitivamente de la red.

Ninguna de las cinco personas entrevistadas llegó a llevar a cabo un proceso de denuncia administrativa o penal. Los motivos, nuevamente, pueden ser varios. Por un lado, Sara

apunta a que, cuando ella recibió los ataques en 2019, **no había tanta conciencia** como en la actualidad sobre la importancia de denunciar el discurso de odio, ni tampoco sobre cómo hacerlo. Aunque a día de hoy haya más concienciación sobre la gravedad de estos discursos, la **falta de tiempo** o la **poca confianza en la administración** son dos de los obstáculos más frecuentes en la denuncia formal.

“Llevo una vida bastante frenética y, si empiezo a denunciar, debería estar denunciando todos los días. Prefiero seguir con mi vida”, defiende Najat Driouech.

Miriam y Najat refieren, sin embargo, haber denunciado una de las veces en organismos públicos como la Oficina por la No-Discriminación en Barcelona o la Oficina d'Igualtat de Tracte i No-Discriminació de la Generalitat de Catalunya. La falta de confianza en la administración mencionada anteriormente se ve reflejada por un lado en la **creencia o convicción que, desde estos espacios, se puede hacer muy poco o casi nada que tenga un resultado tangible** más allá de su registro en estadísticas e informes, en especial si la **persona atacante no puede ser identificada** como es en la mayoría de los casos. “Aun si denuncias a la policía, si no pueden rastrear quien lo está haciendo, o a no ser que sea una amenaza directa muy concreta, tampoco sirve de nada. Lo último que te apetece es perder el tiempo haciendo una denuncia que no llega a ninguna parte”, explica Miriam.

Es innegable que, en el caso de personas en condición de racialización o de personas que se encuentran en situación administrativa irregular, esta **desconfianza también se da hacia los propios cuerpos policiales**. Por un lado, las paradas por perfilación étnica por parte de los cuerpos policiales identifican de manera sistemática a personas en condición de racialización, en cualquier tipo de situación en el ámbito público. Estas identificaciones son motivadas por un **sesgo racista que criminaliza automáticamente las personas no-blancas**. Éstas pueden conllevar violencia a nivel verbal, físico y emocional, además de constituir casos de violencia institucional. Además, pueden resultar en detenciones de personas en situación administrativa irregular que, lejos de gozar de los derechos de cualquier otra persona ciudadana, pueden ser encerradas en los Centros de Internamiento de Extranjeros (CIE) y acabar siendo deportadas. Por otro lado, son asimismo frecuentes las situaciones de descrédito de los discursos de las personas en condición de racialización que atienden a comisaría o los procesos de revictimización.

También existe una **desconfianza hacia el proceso de denuncia** en sí: plazos que se suelen extender años, la poca probabilidad de ganar el caso, el gasto económico y, sobre todo, el desgaste emocional. Este último aspecto se vincula con un motivo que especifica Sara sobre por qué no denunció los ataques, que es el hecho de encontrarse en un país en el que no cuenta con el apoyo y arropo de su familia ni de “su gente”.

“Aunque haya vías legales, piensas, ¿qué va a pasar? Quizás no pasa nada con este continuum de violencia que tienen que sufrir las personas racializadas... entonces, para sobrevivir, se banalizan bastantes situaciones que son jodidas, aunque no deberías estar viviendo esto”, explica Sarahi.

La **aceptación o resignación ante estos ataques de odio**, en especial si se dan de manera reiterada, se convierte en un **mecanismo de autodefensa** ante los mismos y de protección de la propia estabilidad emocional y mental.

La recopilación de estos obstáculos configura parte de la realidad a la que se enfrentan las personas que son atacadas en redes, en especial por ataques racistas, islamófobos y de género. A pesar de estos obstáculos, las personas entrevistadas coinciden con la **necesidad de que existan vías accesibles de denuncia, que todas las personas sean concedoras de sus derechos**, también digitales, independientemente de su condición, y destacan la importancia de denunciar los discursos de odio, independientemente del formato y dentro de las posibilidades de cada persona.

---

## ¿Qué son los derechos digitales?

---

Los derechos digitales son una extensión de los derechos de la Declaración Universal de Derechos Humanos. Aunque no existe consenso internacional, en la teoría tienen por objetivo principal garantizar acceso a Internet, que se haga un uso correcto del mismo y evitar la brecha digital. A nivel europeo existe la declaración sobre los derechos y principios digitales europeos<sup>1</sup>. Publicada en diciembre de 2022, los principios que promueve son:

1. Las personas al centro: las tecnologías digitales deberían proteger los derechos de las personas y reforzar la democracia.
2. Solidaridad e inclusión: la tecnología debería unir a las personas. Todo el mundo debería tener acceso a Internet, habilidades digitales, servicios públicos digitales y condiciones de trabajo justas.
3. Libertad de decisión: las personas deberían beneficiarse de un ambiente en línea justo, seguro de contenido dañino e ilegal, y fortalecerse cuando interactúan con nuevas tecnologías como la inteligencia artificial.
4. Participación: las personas ciudadanas deberían poder participar en los procesos democráticos multinivel y tener control sobre sus datos.
5. Seguridad y protección: todos los usuarios deberían ser fortalecidos y protegidos.
6. Sostenibilidad: los aparatos digitales deberían respaldar la sostenibilidad y la transición ecológica.

La red europea EDRI vela para la defensa de los derechos y libertades en línea. Trabaja para desafiar los actores privados y estatales que abusan de su poder para controlar o manipular al público, abogando por leyes sólidas y aplicadas, informando y movilizándolo a las personas y promoviendo un mercado tecnológico saludable y responsable. Algunos de sus nexos de trabajo son la descolonización de los derechos digitales –de la mano con grupos y comunidades por la justicia social– y la investigación sobre la Inteligencia Artificial (IA).

A nivel nacional, cabe destacar la Asociación AlgoRace, un equipo multidisciplinar que trabaja para la desracialización de la inteligencia artificial. Es decir, para vislumbrar cómo los sistemas de IA incorporan y amplifican el racismo estructural. Para conseguirlo, sus líneas de trabajo se centran en la investigación, la sensibilización y denuncia pública, incidencia política y social. Algunas de sus investigaciones recientes han tratado la tecnología aplicada a los procesos migratorios y cómo afecta a los derechos de las personas que migran, los algoritmos para la vigilancia masiva de personas migrantes y cómo se vincula la IA con el racismo.

---

1 European Commission. (27/09/2023). "European Digital Rights and Principles." <https://digital-strategy.ec.europa.eu/en/policies/digital-principles>

## Seguridad en línea y traslado al plano físico

La división entre el plano físico y el plano digital puede conllevar una sensación de falsa seguridad ante el uso de las redes sociales. Resulta fácil pensar que aquello que se publica en un perfil es de acceso privado y solo personas cercanas accederán a esta información. Sin embargo, **toda la información compartida en redes es de fácil acceso, lo cual facilita la ubicación de una persona en su entorno real habitual**. Por ello, es esencial reconocer el vínculo entre el plano digital y el plano físico y como lo que acontece en un espacio se traslada en el otro.

“Cuando me atacaron en línea, sentí que habían entrado en un espacio propio. A veces piensas que se trata de tu espacio propio en las redes y que es seguro, pero no es así.” - Sara Cuentas

En esta línea, Sara y Daniela refieren la **necesidad de limitar la presencia en redes y de tomar medidas básicas de seguridad** como por ejemplo no interactuar con los mensajes de odio. Convertirse en objetivo de ataques de odio genera una **sensación de miedo que se sustenta en la posibilidad de que estos ataques se materialicen en el plano físico**. “El miedo no es a que te insulten, sino es miedo a que esto se magnifique y tome otro carácter”, explica Daniela. La materialización de esta violencia podría ser física, o también institucional, como el miedo a la retirada de custodia de un hijo a una madre migrante, o el miedo a que la actividad en redes sociales resulte en un obstáculo

## Efectos del discurso de odio

Como consecuencia de los ataques de odio, el **comportamiento de las personas violentadas se ha visto alterado de una forma u otra**, tanto en el plano virtual como en el plano físico. En relación con la seguridad física, los efectos sobre el comportamiento de las personas han sido mínimos, el más destacado es el dejar de compartir los viajes y la ubicación en redes sociales. En las plataformas digitales, en cambio, los cambios en el comportamiento van desde el **uso de las herramientas que brinda Twitter** (privatización de perfil, limitar las interacciones, bloqueo y denuncia de cuentas) hasta la **eliminación de la publicación** que había detonado los ataques o incluso la **eliminación del perfil propio** de manera temporal.

Independientemente de la reacción inmediata, todas las personas entrevistadas coinciden en que estos ataques han conllevado un **distanciamiento de su presencia en la red** social.



Más allá del distanciamiento con la plataforma, los ataques también han tenido una **consecuencia directa en el discurso de las personas violentadas** que se ha traducido en un **silenciamiento parcial** de sus voces. Sarahi admite que es mucho más cautelosa en la hora de publicar y que ha cambiado su forma de hablar: “Siento la necesidad de tener que cuidar las palabras que utilizo para que la gente no me insulte”. Para Najat es inevitable ser más prudente a la hora de compartir publicaciones después de recibir ataques, en especial si se reciben ataques islamófobos independientemente del contenido que se cuelgue. Comparte que le genera **inseguridad** y la desanima:

“A parte de que no me gustan las redes, me generan inseguridad. Yo antes compartía lo que me apetecía. Ahora no. Ahora me lo tengo que mirar bien, pensar y repensar... porque todo el mundo mira lo que publicas, y lo miran con lupa... no me siento con la libertad de antes”, explica.

Esta sensación de tener que cuidar las palabras más que otra persona, para Sarahi refiere a la **hipervigilancia que muchas personas racializadas no-blancas sienten que deben tener**, tanto en el plano físico como en las redes sociales. La sensación de que, por sus condiciones (mujeres, no-blancas, migradas o musulmanas o portadoras de *hijab*) se les está realizando un **examen constante de aquello que publican** o comparten, de que no haya contradicciones o errores en sus discursos.

Estas sensaciones de hipervigilancia y examen constante se vinculan a **la creencia de que una persona en condición de racialización actúa o habla en nombre de toda la comunidad** con la que se la identifica socialmente, siguiendo los criterios de racialización.

Un ejemplo: si una persona blanca roba en un supermercado nadie atribuirá su hurto como algo típico de las personas blancas. En cambio, si una persona negra roba en un supermercado, el castigo social recaerá sobre todas las personas negras, reforzando prejuicios y estigmas racistas.

“Las mujeres que decidimos llevar el *hijab* estamos siendo escrutadas diariamente y públicamente, se nos cuestiona, se nos incapacita. [...] Todo lo que les pasa a las mujeres musulmanas en el mundo es culpa mía, según las redes.” - Najat Driouech

### Redes de apoyo y cuidados

Recibir ataques de odio basados en la mera existencia de una persona y lo que su cuerpo representa en la sociedad tiene un impacto psicológico y emocional que no se debe subestimar, y que no se ve minorizado por el hecho de que se dé a través de redes sociales. Por este motivo, las mujeres entrevistadas hacen referencia a la importancia de **contar con redes de apoyo como estrategia de resistencia**, ya sea en el plano digital o en el plano real.

Sarahi y Míriam apuntan que, cuando han recibido ataques por las redes sociales, ha sido en las mismas plataformas donde han visto **mensajes de confrontación a los atacantes por parte de personas desconocidas**, o con quien no hay un vínculo personal establecido, en su defensa, y mensajes de apoyo por la vivencia sufrida. No obstante, Míriam comparte que en las redes sociales se genera un bucle desagradable de una dinámica ataque-defensa, que no permite expresarte libremente sobre lo que una desea.

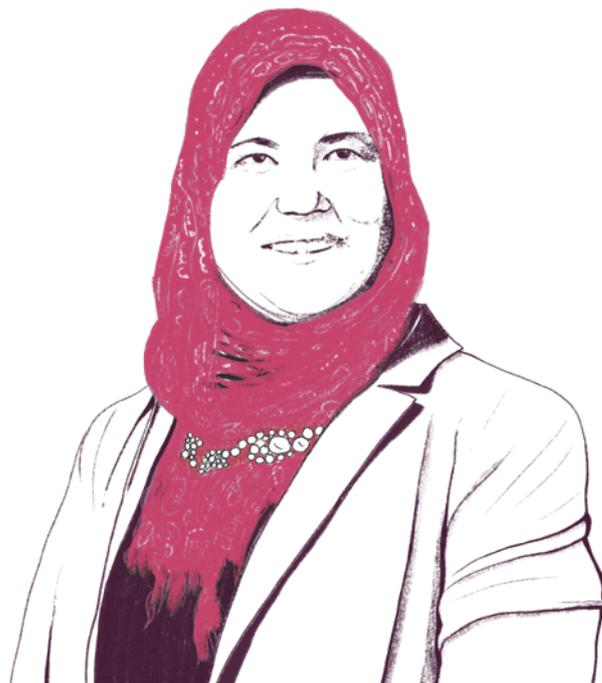
Para Daniela, este apoyo se ha brindado también de forma virtual de **personas conocidas que habían pasado por situaciones similares** y que compartieron con ella consejos, recibiendo también apoyo jurídico sobre cómo compartir contenido en redes de forma segura.

Para Sara, el apoyo de personas que publican o comentan en las redes a su favor no conforma el apoyo que necesita. En cambio, recalca el **apoyo recibido por sus amigas en el plano físico**: “Al final, el apoyo no lo encuentras en la virtualidad, porque es un apoyo virtual (se ríe), es algo que no tocas, es intangible. No es lo mismo que encontrarte con una amiga, abrazarte, reflexionar...” Najat subraya el rescoldo por parte de sus compañeras de trabajo y en su familia y destaca que no espera apoyo por parte de organismos públicos:

## Hacemos frente al discurso del odio

*“Yo he aprendido a no verme como víctima. Es lo que quieren los agresores, que te vayas haciendo más pequeña.”*

**Najat Driouech**



También son estrategias de resistencia comunitaria la **denuncia en colectivo** de un tweet, comentario, o perfil concreto, con la finalidad que tenga un efecto mayor y la medida “punitiva” de la plataforma sea tomada seriamente y de manera más rápida, o el compartir listas de cuentas bloqueadas por parte de activistas populares.

## 6. Recomendaciones

---

Combatir los discursos de odio debe ser una prioridad para las instituciones en el conjunto global de la sociedad. Para ello, se ha realizado una propuesta de recomendaciones dirigidas a Twitter y a las instituciones que pasan por los puntos esenciales de la prevención y sensibilización frente a los discursos de odio, la moderación de las publicaciones e interacciones, el acceso a mecanismos de denuncia con resultados tangibles, el acompañamiento jurídico y psicosocial y la reparación y no-repetición.

**Y a título individual, ¿cuál es la mejor manera de actuar frente a publicaciones de odio en las redes? Te sugerimos echarle un vistazo a [las recomendaciones del informe “Racismo digital y COVID-19”](#).**

### A Twitter:

#### Prevención y sensibilización

- Desarrollar campañas de sensibilización y denuncia frente a los discursos del odio, fomentando el respeto a la diversidad cultural, étnica, religiosa, sexual y de género.
- Llevar a cabo campañas y políticas de disuasión del odio dirigidas a los perfiles de los atacantes.
- Desplegar una serie de recursos pedagógicos, tanto para jóvenes y adolescentes como para el público adulto, que aborden la educación digital y eduquen en el respeto a la diferencia desde la perspectiva de la educación basada en derechos humanos (EBDH) y un enfoque de género.
- Destinar recursos económicos a entidades independientes que tengan la misión de prevenir los discursos de incitación al odio discriminatorio en las redes sociales.

#### Moderación y acción frente a los discursos de odio

- Invertir los recursos y personal necesarios para reactivar y perfeccionar la moderación de contenidos en Twitter con el fin de limitar y erradicar los discursos de odio en la red.
- Priorizar la protección de los perfiles atacados por los discursos de odio ampliando los mecanismos de moderación.
- Ejercer de manera proactiva la aplicación y ejecución de las políticas internas de Twitter contra los discursos de odio, en vez de relegar esta función solamente en los casos de denuncia.
- Asegurar que los protocolos de moderación de la red se apliquen acorde a los derechos humanos desde una interpretación lo más amplia posible en cuanto a su cobertura.

- Garantizar que las herramientas tecnológicas usadas en la moderación no estén basadas en sesgos ideológicos ni perpetúen situaciones de discriminación o persecución política.
- Establecer criterios transparentes en la moderación de las publicaciones en la plataforma.

### **Reparación y no repetición**

- Ofrecer apoyo económico independiente para asistencia legal y acompañamiento psicosocial a las personas receptoras de los discursos de odio.
- Guardar los datos de todos los contenidos eliminados por denuncia de discurso de odio al servicio de las víctimas en caso de que requieran de pruebas para iniciar un proceso judicial.
- Promover programas o cursos de reeducación, concienciación, educación en la diversidad y prevención del odio a aquellas personas responsables de cuentas emisoras de los discursos del odio.

## **A las instituciones:**

### **Prevención y sensibilización**

- Investigar o fomentar el estudio de comportamientos y dinámicas que incitan al odio en el plano digital.
- Realizar campañas de sensibilización y promoción del respeto a la diversidad para la prevención de los discursos del odio.
- Promover políticas públicas destinadas a la concienciación social y fomento de narrativas desde una perspectiva antirracista para disuadir los tópicos, estereotipos y desinformación de carácter racista, machista, LGTBfóbico o discriminatorio.
- Formar y ofrecer recursos al funcionariado del ámbito judicial con el fin de sensibilizar y concienciar de la problemática.
- Apoyar la implicación de las entidades y organizaciones de la sociedad civil en la sensibilización frente los discursos de odio y el diálogo entre las mismas y las instituciones.
- Predicar con el ejemplo y asegurar una actuación institucional responsable comprometida con el respeto a la diversidad, incluyendo los partidos políticos.

### **Acción frente a los discursos de odio**

- Reforzar los mecanismos de detección de los discursos de odio.
- Establecer una unidad especializada en delitos y discursos de odio en redes ajena a las fuerzas y cuerpos de seguridad del estado.
- Fortalecer las redes intra-nacionales entre los mecanismos y entidades detectoras y sancionadoras del discurso del odio.

### **Reparación y no repetición**

- Brindar apoyo a las personas receptoras de los discursos de odio con un programa público y gratuito de acompañamiento jurídico, social y psicológico.

- Simplificar y facilitar el acceso a los mecanismos públicos de denuncia con resultados tangibles y con el fin de eliminar barreras al derecho de reparación.
- Difundir campañas destinadas a la divulgación del derecho a la reparación frente a los discursos de odio y a la importancia de la denuncia de los mismos.
- Impulsar políticas públicas que fomenten el respeto a la diversidad y la disuasión de los mensajes de odio.



**NOVACT**

**novact.org**

**@novact\_**