Remoción de contenidos:
Desigualdad y exclusión del espacio cívico digital

ARTICLE 19

FRIEDRICH NAUMANN STIFTUNG

Este documento fue elaborado con el apoyo de ARTICLE 19, Oficina para México y Centroamérica y la Fundación Friedrich Naumann para la Libertad, Proyecto México. El encauce de la investigación corresponde a Candy Rodríguez, a petición de las organizaciones mencionadas.

Diseño y diagramación:
Isaac Ávila
Ramón Arceo

Ilustraciones:
Sabrina Almanza

Contenido

Introducción 5

Políticas y prácticas de remoción de contenidos 8

Proceso de exclusión y discriminación 12

Censura hacia grupos y poblaciones en situación de mayor vulnerabilidad 13

Relación con otras formas de violencia 18

Efectos adversos en los derechos humanos 18

Consideraciones finales 21

Recomendaciones 22

Introducción

Internet, y en especial las redes sociales, se han convertido en una ampliación del espacio cívico. Es ahí donde las personas y comunidades expresan sus posicionamientos políticos, ideas de índole social e información de interés público que se gesta en este espacio de forma orgánica, o incluso en el terreno físico pero que se traslada al ámbito digital.

Al compartir materiales escritos, visuales y audiovisuales que respaldan su activismo, las personas buscan visibilizar la movilización y las reivindicaciones que empujan e, incluso, exhortar al resto de la sociedad a formar parte de ello. Asimismo, más allá de habilitar un espacio para compartir contenidos y reflexiones, estas tecnologías también han permitido que las voces críticas y disidentes sean replicadas y exponenciadas más allá de sus comunidades cercanas de apoyo y sus espacios geográficos, habilitando las condiciones para convocar actos de protesta y activismo ciudadano.

Sin embargo, es importante recalcar que no todas las personas se encuentran en igualdad de condiciones para expresarse y organizarse libremente en el espacio cívico digital. Personas periodistas, activistas, defensoras de derechos humanos y que conforman colectivas, colectivos, grupos de lucha o resistencia –que han sido históricamente discriminadas por sus orígenes, orientaciones sexuales, políticas o sociales, identidades, o por retar el *status quo*–, enfrentan no solo el rechazo y la exclusión en el espacio cívico físico, sino también la censura y expulsión del ámbito digital.

Como se caracterizará más adelante, las plataformas digitales como Facebook, Twitter, YouTube y TikTok en ocasiones perpetúan y potencializan formas de discriminación a determinados grupos y poblaciones de personas a través de la implementación de sus "términos de servicio" o "normas comunitarias", inclusive, a través del despliegue de algoritmos y tecnologías que invisibilizan sus identidades y sus luchas.[1]

El impacto de lo anterior no es menor, ya que la expresión, la movilización social y la protesta en el espacio cívico digital son acciones esenciales para el funcionamiento de la democracia de nuestra época, ya que permiten impulsar el cotejo de visiones, necesidades y narrativas, y arribar a puntos de encuentro.[2]

No obstante, para que exista una internet que verdaderamente favorezca al desarrollo democrático, es necesario que todas las personas tengan la capacidad de ejercer, en igualdad de condiciones, todos sus derechos y libertades en este entorno.[3] Desafortunadamente lo anterior está lejos de cumplirse ya que, entre muchos otros factores, son las propias plataformas digitales quienes remueven o eliminan contenidos de las personas usuarias, basándose en "términos de servicio" o "normas comunitarias" construidos desde visiones o identidades que, por un lado, desdibujan o se muestran apáticas ante los contextos y realidades que se viven en otras regiones del mundo –diferentes a donde fueron creadas las empresas– y;[4] por el otro lado, porque las reglas establecidas por las plataformas digitales no están del todo alineadas con principios y estándares de derechos humanos.[5]

La remoción o eliminación de contenidos sin apego a estándares internacionales de derechos humanos, según el ex Relator de Libertad de Expresión para las Naciones Unidas, David Kaye, tiene un impacto directo en el derecho a la libertad de expresión, en su dimensión individual y colectiva, ya que

1 Lim, Merlyna y Ghadah Alrasheed, "Beyond a technical bug: Biased algorithms and moderation are censoring activists on social media", *The Conversation*, 16 de mayo de 2021, https://theconversation.com/beyond-a-technical-bug-biased-algorithms-and-moderation-are-censoring-activists-on-social-media-160669

2 Organización de Estados Americanos, "Relatorías de Libertad de Expresión emiten declaración conjunta acerca de Internet" (Comunicado de prensa R50/11), 1 de junio de 2011, https://www.oas.org/es/cidh/expresion/showarticle.asp?artID=848

3 ARTICLE 19,"Derechos humanos son derechos digitales", https://seguridadintegral.articulo19.org/wp-content/uploads/2020/11/art19_2020_infografia-DerechosHumanos_Digitales_V2.pdf

4 Bailey, Matt, "Three Big Discussions We Need to Have ASAP About AI and Social Media Moderation", *PEN America*, 8 de julio de 2020, https://pen.org/three-big-discussions-ai-social-media-moderation/

5 AccessNow, "Protecting Free Expression In The Era Of Online Content Moderation", mayo de 2019, https://www.accessnow.org/cms/assets/uploads/2019/05/AccessNow-Preliminary-Recommendations-On-Content-Moderation-and-Facebooks-Planned-Oversight-Board.pdf y Naciones Unidas, Consejo de Derechos Humanos (CDH), "Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión" (A/HRC/38/35), 6 de abril de 2018, página 23, https://undocs.org/es/A/HRC/38/35

"la información removida no circula, lo que afecta el derecho de las personas a expresarse y difundir sus opiniones e ideas y el derecho de la comunidad a recibir informaciones e ideas de toda índole"[6].

Como también se analizará más adelante, el hecho de que las empresas privadas censuren contenidos o bloqueen a cuentas y perfiles tiene también un impacto directo en el derecho a la igualdad, en la movilización social, en el acceso y en la libre circulación de la información. Además, este tipo de acciones van en detrimento de las reivindicaciones y expresiones sociales, puesto que sus efectos son más abruptos y devastadores cuando éstas se llevan a cabo en contra de grupos o poblaciones en situación de mayor vulnerabilidad.

6 Naciones Unidas, *op. cit.*, página 3.

1

Políticas y prácticas de remoción de contenidos

Las redes sociales han evolucionado de ser habilitadoras de canales de comunicación entre personas, a convertirse en *"gatekeepers"*[7] o "porteras" de los contenidos que se publican. Sus "términos de servicio" o "normas comunitarias" regulan qué y a quiénes se les permite la expresión en sus espacios a través de la moderación de contenidos.[8]

La remoción o eliminación de contenidos se realiza de la siguiente manera: personas que fungen como "moderadoras de contenidos" y algoritmos programados identifican o les son reportados posibles contenidos infractores y, con base en los "términos de servicio", "reglas y políticas" o "normas comunitarias" de las plataformas digitales, determinan qué contenidos pueden permanecer –o no– en línea.

El primer problema de lo anterior es que estas prácticas, muchas veces "incumplen los estándares mínimos de debido proceso y/o limitan de manera indebida el acceso a contenidos o su difusión"[9].

El segundo problema refiere a la dependencia que estas plataformas han desarrollado hacia las herramientas automatizadas para remover contenidos, ya que dicha tecnología "no puede leer los matices del habla –como los humanos–, y para algunos idiomas apenas y funciona"[10]; por lo que "el uso de la automatización ha dado lugar a numerosas eliminaciones indebidas [de contenidos]"[11].

Agudizando el problema, las personas moderadoras de contenido se enfrentan a condiciones laborales precarizadas que no facilitan su labor. La carga de trabajo es excesiva, no reciben atención psicológica para lidiar con los contenidos a los que son expuestos –muchas veces violentos y atroces–, y deben cumplir con un estándar de número de revisiones por día.[12] Además, no cuentan ni son provistas con el contexto adecuado para realizar evaluaciones pertinentes sobre los contenidos que revisan.[13] Por lo tanto, es probable que se vean orilladas a dejar que prejuicios morales o estándares personales los guíen –consciente o inconscientemente– a lo que eliminan de internet.[14]

Más recientemente, durante 2020 la figura de "moderadoras de contenido" se limitó por la pandemia que causó el virus SARS–CoV–2. Muchas personas con este rol tuvieron que trabajar de manera remota[15], en ocasiones enfrentándose a contextos no idóneos para desempeñar su labor. Aunado a ello, en un esfuerzo por atajar los contenidos de desinformación que abundaron sobre conspiraciones del origen del virus, supuestos tratamientos y curas para la COVID-19 y otras informaciones dañinas para la salud, las plataformas de redes sociales aumentaron el uso de algoritmos automatizados[16] para restringir contenidos en sus espacios. Al final, con estas circunstancias se aplican criterios mucho más generales en

7 En términos del poder que han consolidado algunas empresas para"controlar el acceso" a servicios críticos en línea que permiten llegar a una gran categoría de personas usuarias. - Mani, Vivek, "Taming Gatekeepers – But Which Ones?", *The National Law Review*, 25 de febrero de 2021, volumen XI, número 56, https://www.natlawreview.com/article/taming-gatekeepers-which-ones

8 ARTICLE 19, "Side-stepping rights: Regulating speech by contract", 2018, https://www.article19.org/wp-content/uploads/2018/06/Regulating-speech-by-contract-WEB-v2.pdf

9 Organización de Estados Americanos, "Declaración Conjunta Sobre Libertad De Expresión Y "Noticias Falsas" ("Fake News"), Desinformación Y Propaganda", 2017, http://www.oas.org/es/cidh/expresion/showarticle.asp?artID=1056&lID=2

10 Electronic Frontier Foundation, "Automated Moderation Must be Temporary, Transparent and Easily Appealable", 2 de abril de 2020, https://www.eff.org/deeplinks/2020/04/automated-moderation-must-be-temporary-transparent-and-easily-appealable

11 *Ibidem*

12 Solon, Olivia, "Underpaid and overburdened: the life of a Facebook moderator", 25 de mayo de 2017, https://www.theguardian.com/news/2017/may/25/facebook-moderator-underpaid-overburdened-extreme-content

13 *Ibidem*

14 Díaz, Ángel, y Laura Hecht-Felella, "Double Standards in Social Media Content Moderation", *Brennan Center for Justice at New York University School of Law*, 4 de agosto de 2021, página 11, https://www.brennancenter.org/sites/default/files/2021-08/Double_Standards_Content_Moderation.pdf

15 Godoy, Juan Diego, "YouTube relega a las máquinas y recupera a los moderadores humanos para filtrar el contenido dañino", *El País*, 21 de septiembre de 2020, https://elpais.com/tecnologia/2020-09-21/youtube-relega-a-las-maquinas-y-recupera-a-los-moderadores-humanos-para-filtrar-el-contenido-danino.html

16 *Ibidem*

la remoción de contenidos, derivando en la censura masiva de contenidos legítimos.[17]

Así, debido a las reglas determinadas por las plataformas digitales o por las medidas que se han tomado –sobre todo– en el contexto de la pandemia, la diversidad de los contenidos e información en línea ha disminuido considerablemente y, en determinados casos, ha puesto en situación de desigualdad a las personas usuarias que publican y acceden a información a través de plataformas sociales[18].

Las plataformas YouTube, Facebook, Twitter y TikTok son las redes sociales más utilizadas por las y los mexicanos,[19] por lo que el presente documento se acotará a proveer información sobre dichas plataformas con miras a abrir la discusión sobre la magnitud del impacto de la remoción de contenidos en los derechos humanos.

- **Facebook:** A pesar de ser "parte de una iniciativa global que ofrece a las empresas de Internet un marco para aplicar los principios de derechos humanos"[20], la compañía ha señalado que elimina contenidos que considera que ponen "en peligro física o financieramente, que [intimidan] a las personas mediante un lenguaje de odio o que [tienen] como objetivo lucrar engañando a las personas que usan Facebook".[21]

- **TikTok:** La plataforma ha declarado que tiene como objetivo el fomentar la creatividad de las personas usuarias y generar espacios en donde una comunidad global pueda crear contenido, compartir, descubrir el mundo y crear relaciones con personas de todo el planeta.[22] No obstante, describen que eliminarán "cualquier contenido (incluido vídeo, audio, transmisiones en vivo, imágenes, comentarios y texto) que infrinja las Normas de la Comunidad. Las sanciones consistirán en la suspensión o prohibición [de] las cuentas y/o los dispositivos desde los cuales se hayan producido incumplimientos graves o repetidos"[23].

- **Twitter:** La compañía enfatiza que busca promover la conversación pública a través de su plataforma[24]. Sin embargo, que "el acoso, los discursos de odio y otras actitudes similares no incentivan a las personas a expresarse y disminuyen la conversación pública"[25]; por lo que la plataforma podrá determinar la eliminación del contenido si determina que "el Tweet incumplió las Reglas de Twitter", solicitando "al infractor que lo elimine antes de poder volver a twittear".[26]

- **YouTube:** La red social indica que permite "subir" o "cargar" cualquier tipo de contenido a su espacio, siempre y cuando no incumpla con sus "términos de servicio"[27]. Las reglas de esta empresa "constituyen una especie de contrato entre Google y las y los usuarios de cualquiera de sus aplicaciones",[28] en este caso, YouTube. En este contrato viene información sobre el desarrollo de los servicios que ofrecen la plataforma y las reglas, los derechos legales y de autor que se deben cumplir las personas usuarias; así como los contenidos que están restringidos: como las estafas, el *spam*, la suplantación de identidad, la pornografía o desnudez, el suicidio o las autolesiones, así como el 'lenguaje vulgar'.[29]

Aunque las redes sociales son espacios de expresión, libertad y debate público, a su vez también justifican y aplican una mordaza a expresiones y contenidos desde un punto de vista –muchas veces– arbitrario y que no está alineado con el derecho internacional de derechos humanos.

Para luchar contra esta situación, es cierto que en estas plataformas existen posibilidades de apelación ante la eliminación de contenidos, cuentas y perfiles de personas usuarias, pero, como se verá en la siguiente sección, las personas consultadas en la presente investigación señalaron que estos re-

17 Greenspan, Rachel E. y Kat Tenbarge,"YouTubers' channels and videos are being mistakenly deleted for debunking COVID-19 conspiracy theories", *Insider,* 26 de septiembre de 2020, https://www.insider.com/youtube-demonetized-covid-19-disinformation-moderation-automation-bots-strikes-2020-9

18 Guerrero, Carlos, "Políticas de la moderación de contenidos", *Hiperderecho,* 25 de febrero de 2020, https://hiperderecho.org/2020/02/politicas-de-moderacion-de-contenidos/

19 En México, 69% del total de la población son personas usuarias de las redes sociales. Considerando el total de usuarias y usuarios de internet entre los 16 y 64 años, las diez redes sociales más utilizadas en México son, por orden de popularidad: YouTube, Facebook, WhatsApp, Facebook Messenger, Instagram, Twitter, Pinterest, Linkedln, SnapChat y Skype. - Hootsuite y We Are Social, "Digital 2020: Mexico", 11 de febrero de 2020, https://datareportal.com/reports/digital-2020-mexico

20 Facebook, "Hard questions: Where do we draw the line on free expression?", 9 de agosto de 2018, https://about.fb.com/news/2018/08/hard-questions-free-expression/

21 *Ibidem*

22 TikTok, "Normas de la comunidad", diciembre de 2020, https://www.tiktok.com/community-guidelines?lang=es#29

23 *Ibidem*

24 Twitter, "Las Reglas de Twitter", https://help.twitter.com/es/rules-and-policies/twitter-rules

25 *Ibidem*

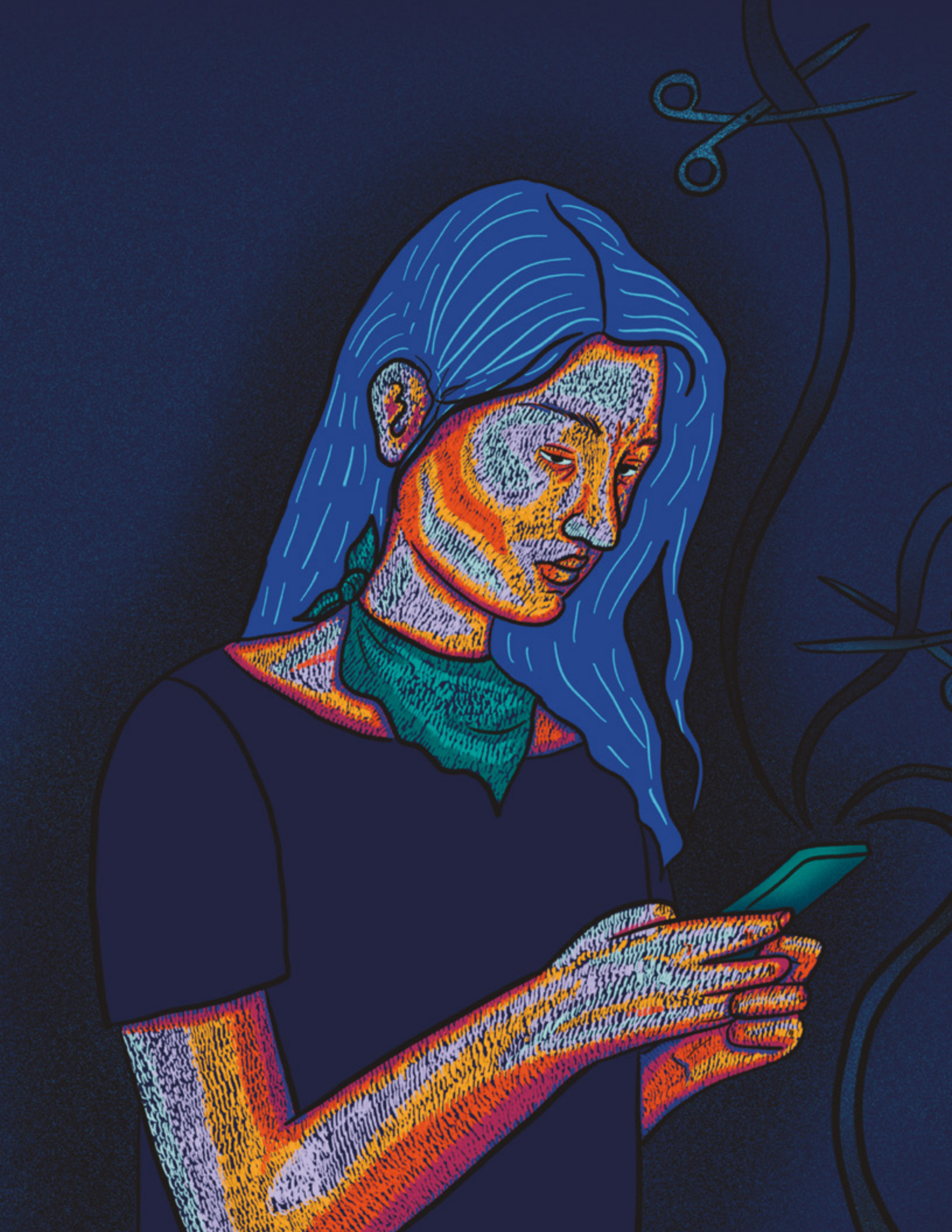
26 ARTICLE 19 e Indela, "Guía sobre Remoción de Contenido Sobre las Reglas y Políticas de Twitter," página 5, https://seguridadintegral.articulo19.org/wp-content/uploads/2020/10/Article19_ManualTwitter_Web.pdf

27 ARTICLE 19 e Indela, "Guía de remoción de contenido sobre las políticas de Google", página 4, https://seguridadintegral.articulo19.org/wp-content/uploads/2021/03/GUIA-REMOCION-DE-CONTENIDO-GOOGLE.pdf

28 *Ibidem*

29 *Ibidem*

portes generalmente se atienden de forma automatizada, descontextualizada y aletargada. En caso de lograr apelar con éxito y revertir la decisión de la plataforma digital, la restitución de los contenidos llega a resultar fuera de tiempo o de la coyuntura social donde tal contenido fue necesitado, compartido y buscó ser difundido.



2

Proceso de exclusión y discriminación

Para vislumbrar las afectaciones más puntuales de la remoción de contenidos en grupos o poblaciones en situación de mayor vulnerabilidad, desde ARTICLE 19 se realizó una investigación desplegada a través de una encuesta[30] y una serie de grupos de enfoque[31] con personas en México que se auto–adscriben a: (1) la comunidad de la diversidad sexogenérica (LGBTTTIQ+); (2) movimientos de mujeres organizadas e iniciativas feministas; (3) la población afrodescendiente e/o indígena; y (4) grupos en defensa de la tierra, territorio y cultura.

En estos ejercicios participaron un total de 63 personas de 14 estados del país. El principal objetivo fue generar una aproximación sobre cómo las políticas y prácticas de moderación de contenidos de las plataformas digitales impactan específicamente en las identidades o causas que las personas –de poblaciones o grupos en situación de mayor vulnerabilidad– reivindican, visibilizan, defienden y/o empujan. También, en cómo afectan las expresiones y movilizaciones digitales de estas personas de forma diferenciada. Y, por último, en qué manera se relacionan con el ejercicio del derecho a la igualdad y el derecho a libertad de expresión.

A continuación, se comparten algunos de los principales resultados de la investigación:

- El 88% de las personas participantes afirmaron que el contenido que ha sido censurado de sus perfiles y cuentas de redes sociales –a través de la remoción de contenidos– está directamente vinculado con la reivindicación de sus identidades, de sus derechos humanos o con su activismo político.

- Las redes sociales con más casos de remoción de contenidos –hacia la población de personas consultadas en la investigación– fueron, por orden de reincidencias, Facebook, Twitter, Youtube y Tik Tok.

- Las imágenes[32] fueron los contenidos más eliminados (correspondiendo al 32.5% de las publicaciones reportadas como suprimidas), el texto o expresiones escritas fue la segunda categoría de contenido más eliminado (31.25%), siguiendo los perfiles de las personas usuarias (20%) y, por último, los videos (16.25%).

- El 61% de las personas encuestadas confirmaron que, tras la eliminación de sus publicaciones, la plataforma no les dio ninguna razón por la cual su contenido fue censurado. Solamente en 16 casos se les ofreció una explicación.

- La expresión individual y la movilización social colectiva han estado influenciadas por la pandemia por COVID-19. De acuerdo a la gran mayoría de las personas consultadas, las condiciones de confinamiento han incidido drásticamente en el aumento del uso de las TIC para el desarrollo de sus acciones de activismo. En palabras de una persona que se auto–adscribe a la población indígena, *"los medios tradicionales [como la radio] tienen su audiencia, sobre todo la gente mayor; pero también hay una población fuerte que está en redes sociales, y queremos hablarle a esa gente, queremos llegar a ellos. La mayoría de los seguidores que tenemos llegaron el año pasado [2020, cuando comenzaron las indicaciones de contención sanitaria] y siguen llegando ahora. Pero el año pasado fue un parteaguas en el alcance que tenemos".*

- La brecha digital se confirma como una condición que imposibilita el ejercicio de los derechos humanos, sobre todo para aquellas personas activistas que ya experimentan diversas formas de marginación. El primer problema al que se enfrentan es trascender los obstáculos para conectarse materialmente a internet. *"No tenemos buenos equipos, buen internet o buena conexión. Todo eso nos genera ciertas dificultades para difundir nuestro trabajo, para 'subir' cosas"*, menciona una persona indígena activista. Según también lo recuperado en la investigación, el segundo problema al que se enfrentan –como se describirá más adelante– es que sus contenidos permanezcan

30 Desplegada entre los meses de mayo y julio de 2021.

31 Realizados durante julio y agosto de 2021.

32 Esta categoría se refiere a imágenes visuales de toda índole, como retratos, fotografías, memes, ilustraciones o infografías, por ejemplo.

en línea y no sean indebidamente censurados por las plataformas digitales a través de la remoción de contenidos.

Censura hacia grupos y poblaciones en situación de mayor vulnerabilidad

En México, las personas que cuentan con conexión a internet han encontrado en las plataformas digitales una manera de expandir y difundir sus intereses, trabajos y activismos. *"Mucho de lo que hacemos en YouTube, por ejemplo, los compañeros de la radio lo retransmiten. Yo personalmente hago muchos programas en vivo en YouTube, entrevistas con personas que [se expresan en] zapoteco o alguna lengua indígena; y los compañeros de la radio bajan el audio, lo editan un poco y ya lo transmiten [por este medio]. En el colectivo nos gustaría circular en las redes sociales el trabajo cinematográfico que pretendemos hacer, subirlo a YouTube, a Facebook e Instagram"*, compartió una persona que se identifica como activista en defensa de los derechos culturales de la población zapoteca. Sin embargo, esfuerzos e intenciones similares al caso compartido son indebidamente coartados por las políticas y prácticas de moderación de contenidos de las empresas de redes sociales.

En relación con los testimonios de las personas que participaron en la investigación, se identifica que el idioma, el lenguaje y la falta de entendimiento del contexto son factores que aumentan la probabilidad de que las plataformas digitales eliminen contenidos de grupos o poblaciones en situación de mayor vulnerabilidad.

Discriminación por idioma

De acuerdo a información proveniente de fuentes oficiales, en México se hablan y escriben 69 idiomas: 68 indígenas y el español.[33] A pesar de estos datos, el español parece haber sido elegido por las plataformas digitales como el idioma predilecto de ameritar reconocimiento para cómo se tomarán las decisiones de moderación de contenidos en el país.

Las implicaciones de lo anterior no son menores, ya que los idiomas de "alto recurso"[34], como el español, predominan en los algoritmos[35] que filtran y remueven contenidos de internet.[36] Por ende, es más probable que las plataformas bloqueen o censuren los contenidos de comunidades específicas, particularmente de aquellas que se comunican en idiomas autóctonas y no coloniales que no califican como de "alto recurso", debido a que la automatización de moderación de contenidos no es capaz de procesar sus formas de expresión y comunicación.[37] Esto conlleva a una censura desproporcionada hacia ciertos grupos o poblaciones.

Una agrupación oaxaqueña que utiliza las redes sociales para promover, difundir y preservar la lengua zapoteca –dentro y fuera de su comunidad– posiblemente ha experimentado discriminación por el idioma en el cual realiza su activismo, de acuerdo a la perspectiva de uno de sus integrantes. En sus propias palabras relató el siguiente suceso: *"Tuiteamos unas palabras en zapoteco y pusimos [un] audio. Grabamos el audio para decir 'se pronuncia de esta manera'. El plan era que la gente escuchara [una] palabra, y que supiera cómo se dice, cómo se pronuncia. [Las palabras eran] lluvia, agua y sol. Y de repente el contenido desapareció sin aviso. Y no recibí ninguna notificación o si infringí las normas comunitarias o qué pasó, simplemente lo 'quitaron'. No me dio tiempo de buscarlo. Pero fue muy extraño, porque yo pensé que era porque el algoritmo no entendía zapoteco, o porque lo asociaba a otra lengua para lo que podía ser algo ofensivo"*, relató.

No obstante, su sospecha de ser discriminadas y discriminados por el idioma en el que se expresa la agrupación no se basa en esa única incidencia, ya que en Facebook su contenido ha sido también removido. Según compartió: *"Me ha pasado que, de repente, cuando comparto algo [de material en zapoteco], me dice [Facebook] que 'no se puede publicar, porque viola las normas de la comunidad'; pero no me dice qué nor-*

33 Secretaría de Cultura del Gobierno de México, "¿Sabías que en México hay 68 lenguas indígenas, además del español?", 21 de febrero de 2018, https://www.gob.mx/cultura/articulos/lenguas-indigenas?idiom=es

34 Los idiomas de "alto recurso", o los *high-resource languages* -en inglés-, son aquellos para los que existe una gran cantidad de recursos de entrenamiento de datos para la inteligencia artificial, lo que facilita el entrenamiento de modelos de aprendizaje automático para reconocer esos idiomas. - Center for Democracy & Technology, "Mixed messages? The limits of automated social media content analysis", noviembre de 2017, https://cdt.org/wp-content/uploads/2017/11/Mixed-Messages-Paper.pdf

35 Las herramientas que se usan para la moderación de contenidos y están relacionadas con el procesamiento del lenguaje (PNL) están diseñadas para predecir qué tipo de contenido se enmarca en determinados discursos que no cumplan con las normas comunitarias de la plataforma. Regularmente las herramientas PNL están sesgadas porque están configuradas para comprender contextos muy particulares, y difícilmente pueden descifrar otros idiomas, contextos, o las connotaciones con las se compartieron los materiales. - *Ibidem.*

36 Hirschberg, Julia y Christopher D. Manning, "Advances in Natural Language Processing", 12 de mayo de 2016, https://cs224d.stanford.edu/papers/advances.pdf

37 *Ibidem*

mas. No me dice exactamente por qué [se da la censura] y qué parte del contenido viola las normas, y qué tipo de normas son las que viola: si es por raza, por religión, por sexo, por clase social, o lo que sea. No te dice [Facebook], simplemente dice 'fue bloqueado porque hay algún tipo de problema'".

En 2019, antes de la pandemia por COVID-19 y que las plataformas digitales dependieran más de las herramientas automatizadas para la moderación de contenidos, Facebook reportó que su fuerza laboral de personas moderadoras hablaba alrededor de 50 idiomas, y que los algoritmos –en ese momento– funcionaban en aproximadamente 30 de ellos.[38] Tomando el ejemplo de esta red social, y comparándolo con el hecho de que en el mundo se hablan aproximadamente 7000 idiomas distintos (donde 600 cuentan con más de 100.000 hablantes),[39] se puede razonar por qué y de qué manera el idioma funge como un filtro de discriminación y un factor de exclusión del entorno cívico digital.

Si las personas y los algoritmos que revisan los contenidos en las redes sociales desconocen el mero idioma en que individuos, grupos y poblaciones se comunican, es consecuente que se generarán eliminaciones ilegítimas y desproporcionadas de sus publicaciones, por el solo hecho de haberse atrevido a reivindicar sus identidades y ejercer sus tradiciones habladas.

Discriminación por el lenguaje

Trascendiendo la discriminación por el idioma, según los hallazgos de la investigación desplegada, el lenguaje –es decir, la jerga, las palabras y los modismos que se utilizan al comunicarse– también es factor para la remoción de contenidos por parte de las plataformas digitales.

En ese sentido, los primeros relatos de quienes participaron en la investigación provienen de una persona que forma parte de la comunidad de la diversidad sexogénerica. Según su experiencia, es usual que la comunidad LGBTTTIQ+ resignifique –en su hablar– la implicación de palabras que han sido históricamente utilizadas para señalarles y agredirles; como 'joto' y 'maricón', por ejemplo, sin que ello signifique ofensa alguna o sea una expresión de odio. No obstante, esta persona ha sufrido remociones de contenidos por el lenguaje que utiliza. La primera ocasión en que ocurrió fue narrada como sigue: *"Subí una foto de una marcha en donde yo aparezco con una bandera del arcoíris atrás y escribo 'siempre joto, nunca injoto'. Dos segundos después llega la notificación [de Facebook] y la imagen está bloqueada. [...] Me parece que el asunto del contexto tiene mucho que ver [para determinar si cierta publicación debe o no borrarse]".* La segunda ocasión, sin embargo, tuvo afectaciones más desproporcionadas: *"Simplemente, por utilizar la palabra 'homosexuales' [en la red social] me bloquearon 30 días. No había foto. Decía [el texto] 'buenos días, homosexuales'".*

Otro caso emblemático proviene de mujeres que buscan visibilizar las violencias de género a las que están sujetas en el país. Una de las experiencias compartidas, en palabras de una activista que sufrió la remoción de contenidos, fue la siguiente: *"Hace unos meses me censuraron en Facebook algunas cosas. Me molesté y apelé, pero aun así Facebook seguía diciendo que yo [...] había utilizado lenguaje de odio, cuando no fue así. Yo solo expresé algo en contra de los hombres machistas".* Según aseveró, las expresiones que utilizó no incitaban –en ninguna medida– a generar daño hacia los hombres; solo buscó denunciar a aquellos que violentan a las mujeres en un país feminicida, como México.

Las condiciones que habilitan la censura de publicaciones en las redes sociales –por razón del lenguaje– son diversas, pero en lo general esto se debe a que gran parte de lo que se considera como 'prohibido' por las redes sociales no considera el contexto social que le brinda significado a las expresiones en sí.[40] Asimismo, los algoritmos y las personas moderadoras de contenido generalmente no conocen los términos e intencionalidades de los contenidos que están revisando, a pesar de que conozcan el idioma.[41]

Por las experiencias aquí expuestas se puede dilucidar, por un lado, que "un término que se considera un insulto –cuando se dirige a un grupo marginado– puede usarse de manera neutral o positiva entre los miembros del grupo objetivo".[42] Por el otro lado, las denuncias públicas que utilizan ciertas palabras o términos 'prohibidos' por las plataformas digitales no siempre son expresiones que busquen la persecución y linchamiento de ciertos actores en específico, o que terminen por constituir lenguajes no protegidos por el derecho humano a la libertad de expresión.

Así, el lenguaje no comprendido ni contextualizado termina siendo eliminado por las redes sociales, y las expresiones de las comunidades marginadas terminan por ser diluidas del entorno digital.

38 Fick, Maggie y Paresh Dave, "Facebook's flood of languages leave it struggling to monitor content", Reuters, 23 de abril de 2019, https://www.reuters.com/article/us-facebook-languages-insight-idUSKCN1RZ0DW

39 Romero, Sarah, "¿Cuántos idiomas se hablan en el mundo?", *Muy Interesante,* 10 de octubre de 2019, https://www.muyinteresante.es/cultura/arte-cultura/articulo/icuantos-idiomas-se-hablan-en-el-mundo

40 Ghaffary, Shirin, "The algorithms that detect hate speech online are biased against black people", Vox, 15 de agosto de 2019, https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter

41 *Ibidem*

42 Díaz, Ángel, y Laura Hecht-Felella, *op. cit.,* página 11.


NUNCA INJ

Discriminación por falta de entendimiento del contexto

Si bien las empresas indican que reconocen la importancia del contexto y la evaluación de patrones de comportamiento para moderar contenidos en sus espacios,[43] las experiencias de las personas que participaron en la investigación sugieren lo contrario.

Las colectivas feministas que comparten contenido relacionado con derechos sexuales y reproductivos de las mujeres consideran haber sufrido la remoción de contenidos por violar normas comunitarias ajenas a los contextos en los que las expresiones se desarrollan. En cuanto a la experiencia de una activista de este ámbito, ella comparte lo siguiente: *"Recientemente hice un programa de televisión que hablaba sobre menstruación. Lo pasamos por el canal [de YouTube] y lo retransmitimos a través de las redes sociales Facebook y Twitter. Los dos programas se bajaron porque hablábamos de menstruación. De acuerdo a las reglas de Facebook y YouTube no puedes hablar tan tácitamente del cuerpo humano. [Al hablar de menstruación en nuestro programa] nunca fue de manera lasciva ni malintencionada. Solo dijimos la palabra 'menstruación' y –ese tipo de expresiones naturales, incluso otras expresiones de arte donde se muestran los cuerpos– incurren en una situación de estar contra las normas comunitarias de las redes sociales. Lo que hicimos fue un análisis sociocultural de la menstruación y una mirada patriarcal de ella".*

En la experiencia de otra colectiva de mujeres existe la apreciación de que la remoción de contenidos es aún más dañina cuando se da en una coyuntura donde es más crucial que la información y las publicaciones permanezcan en línea. El caso desde el cual surge dicha apreciación se transcribe de acuerdo a lo señalado por una de sus integrantes: *"La persona que en ese momento [2019] era alcalde[44] en Yucatán sacó un video en YouTube criminalizando la protesta feminista. Esa misma persona durante estas elecciones [2021] se postuló a gobernador del estado de Yucatán y ganó. Antes de la veda electoral subimos un post donde pusimos un screenshot y un meme recordando cómo en 2019 él criminalizó la protesta feminista. Consecuencia a ello nos 'puso' una demanda y se abrió una carpeta de investigación. [Nuestra intención al publicar el meme fue hacer] un recordatorio para saber por quién votamos, y ese meme nos lo bajó Facebook. De todas maneras [aunque Facebook eliminó la publicación] lo volvimos a subir, y Facebook nos mandó una advertencia que decía que no podíamos publicar ese tipo de contenido".*

Los algoritmos no entienden el contexto detrás de muchas o la mayoría de las publicaciones fuera de Estados Unidos –país donde son programados por personas que corresponden a realidades distintas a las de México–[45]. Ello, en combinación con la falta de capacitación, sensibilidad y de condiciones adecuadas de trabajo para los revisores humanos, las publicaciones son dadas de baja.

En sí, las decisiones de remoción de contenido que no incorporan los diversos contextos sociales, culturales y coyunturales se vuelven sumamente problemáticas, ya que toman decisiones que vulneran los derechos humanos de todas las personas, pero más aún de aquellas de grupos y poblaciones en situación de mayor vulnerabilidad. Por esa razón, es de suma importancia que este tipo de decisiones tomen en cuenta los contextos, además del género, la raza, la etnia y el territorio donde las expresiones tienen lugar.[46]

43 *Ibid*, página 8.

44 Las plataformas de redes sociales, además de no incorporar de manera suficiente los análisis de contexto para determinar la pertinencia de una remoción de contenido, tampoco consideran cabalmente el marco del derecho internacional de los derechos humanos en materia de libertad de expresión. Por ejemplo, en este determinado caso, en el Sistema Interamericano de Derechos Humanos se ha establecido que los límites de crítica son más amplios cuando ésta se refiere a personas que, por dedicarse a actividades públicas o por el rol que desempeñan en una sociedad democrática, están expuestas -y sujetas- a un control más riguroso de sus actividades y manifestaciones. - Organización de Estados Americanos, "3 - Capítulo II – Evaluación sobre el Estado de la Libertad de Expresión en el Hemisferio", http://www.oas.org/es/cidh/expresion/showarticle.asp?artID=610&lID=2

45 *Ibid*, página 18.

46 Dirk Hovy, "Demographic Factors Improve Classification Performance", *Center for Language Technology*, julio de 2015, https://aclanthology.org/P15-1073.pdf


Reivindicación de derechos sexuales y reproductivos de las mujeres van en contra de las normas comunitarias.

Relación con otras formas de violencia

La vulnerabilidad, la inseguridad y la desconfianza que viven las personas en el espacio cívico afecta cómo se desenvuelven y expresan a través de canales digitales. La remoción de contenidos no ha sido la única forma de exclusión –a través de las TIC– que han experimentado los grupos y poblaciones que participaron en la presente investigación.

Según las experiencias de las personas consultadas, los casos de *doxing*[47] que han sufrido, así como las campañas coordinadas de ataques, amenazas y hostigamientos a los que se han enfrentado, han afectado la manera en la cual comunican y difunden sus identidades, causas e intereses –tanto de forma individual como colectiva– en el espacio digital.

Más aún, a medida en que la crisis sanitaria por COVID-19 hizo necesario replegarse dentro de los domicilios y trasladarse a internet para participar en la vida pública del país, determinadas personas –por sus características, identidades y adscripciones– enfrentan mayor exposición respecto a sus causas, pero también mayor vulnerabilidad al enfrentar la apatía y la discriminación en línea por otras personas usuarias de las tecnologías. Un ejemplo de este sentir proviene de un grupo de mujeres organizadas, las cuales señalan que *"la violencia hacia [su] colectiva ha aumentado mucho a la par de [su] número de seguidores"* en las redes sociales.

La censura y violencia en internet son síntomas del antagonismo y marginación que ya persiste en el entorno físico, en vez de ser conductas aisladas y aleatorias. De los testimonios que respaldan lo anterior destaca la vivencia de una persona perteneciente a la comunidad de la diversidad sexogenérica, la cual compartió lo siguiente: *"en febrero me cerraron tres veces mis cuentas [Twitter, Facebook e Instagram]. Me sentí vulnerable porque, aparte de eliminar mis cuentas, me llegaban amenazas no solo a mí, sino también a mi familia. Yo no sé ni cómo ni de dónde sacaron mi dirección, de dónde sacaron el nombre de mi mamá... pero me llegaban amenazas [...] En ese punto de mi vida me sentí muy vulnerado"*.

La publicación por la cual experimentó la remoción de contenido y los ataques en las redes sociales fue una donde expresaba apoyo hacia una compañera suya, la cual estaba siendo agredida, acosada y discriminada selectivamente por parte de un grupo transfóbico. Como ya se ha dicho, el contexto importa, por lo que es significativo recalcar que en México persisten los crímenes de odio en contra de las personas LGBTTTIQ+,[48] y es probable que los ataques que esta persona enfrentó, así como los que su compañera vivió, fueron selectivos por las características que les distinguen.

De esta forma, se concluye que, si cierto grupo o población en situación de mayor vulnerabilidad no es silenciado directamente por parte de la plataforma digital, es posible que lo sea tras ser víctima de violencia por parte de otras personas usuarias de las TIC.

Efectos adversos en los derechos humanos

La suma de violencias que enfrentan los grupos y poblaciones en situación de mayor vulnerabilidad, incluida la remoción de contenidos, contribuyen a generar un clima de exclusión, censura y apatía social en el entorno digital. Lo anterior, aun cuando la jurisprudencia interamericana ha señalado expresamente que los discursos que expresan elementos esenciales de la identidad y dignidad –como los impulsados por las personas consultadas en la presente investigación– están especialmente protegidos por la libertad de expresión.[49]

"Nos han violentado muchísimo, sobre todo en Facebook. [M]e siento muy desgastada por esto. A consecuencia de esto me he desconectado [de las redes sociales]", expresó una integrante de una colectiva feminista. Otra mujer, también activista por causas feministas afirmó: *"Cuando uso redes sociales para compartir mi causa regularmente me siento vulnerable, me siento señalada. Todo pasa por un doble filtro conmigo misma cuando quiero compartir o decir algo... porque pienso en las consecuencias que puede tener"*.

Continuando la tendencia de lo compartido por dichas mujeres, una persona de la comunidad LGBTTTIQ+ compartió lo que

47 Práctica que consiste en la recopilación y difusión de información personal, a menudo de carácter privado o sensible, sin el consentimiento de la persona afectada. - Red en Defensa de los Derechos Digitales, "El doxing: ¿cuáles son los límites de lo público?", 30 de mayo de 2019, https://r3d.mx/2019/05/30/el-doxing-cuales-son-los-limites-de-lo-publico/

48 El Financiero, "Estos son los estados con más crímenes de odio contra la comunidad LGBTTTI+ en México", 17 de mayo de 2021, https://www.elfinanciero.com.mx/nacional/2021/05/17/estos-son-los-estados-con-mas-crimenes-de-odio-contra-la-comunidad-lgbttti-en-mexico/

49 Relatoría Especial para la Libertad de Expresión de la Comisión Interamericana de Derechos Humanos, "Marco jurídico interamericano sobre el derecho a la libertad de expresión", 30 de diciembre de 2009, párrafos 53-56, páginas 19 y 20, https://www.oas.org/es/cidh/expresion/docs/publicaciones/MARCO%20JURIDICO%20INTERAMERICANO%20DEL%20DERECHO%20A%20LA%20LIBERTAD%20DE%20EXPRESION%20ESP%20FINAL%20portada.doc.pdf

sigue: "*Las redes sociales me provocan ansiedad [y] estrés de unos años para acá. [Me] autocensuro mucho porque a veces me limito a publicar ciertas cosas o imágenes porque, en mi caso, no tengo la piel gruesa y [la violencia] me afecta*". Otra persona, también de la comunidad de la diversidad sexogenérica, coincide con la percepción anterior: "*Me siento vulnerable porque llegamos a espacios [las redes sociales] que no son tan seguros. [Me] ha tocado autocensurarme para no exponerme a la remoción de contenidos*".

Coincidiendo, las sensaciones descritas son a su vez compartidas por una persona activista por la defensa de la cultura, que mencionó lo siguiente: "*tengo desconfianza de las redes y soy cerrada en ese sentido; solo las uso para dar voz a los proyectos. En mis programas hablo sobre expresiones artísticas y temas culturales, y soy atacada. [P]ara estar en redes hay que tener la coraza dura para seguir adelante con el mensaje*".

Así, ambos factores, la violencia –en sus distintas versiones– y la remoción de contenidos generan un efecto inhibidor en la libertad de expresión. Ocasionan que las personas se autocensuren para no ser señaladas y asediadas, y que las personas sean silenciadas por lo que sí llegan a compartir.[50] A su vez, los daños derivados tienen consecuencias sistémicas para una democracia: socavan el avance de la equidad y la inclusión y paralizan la libertad de expresión.[51] "Esta disparidad sienta las bases para un ecosistema en línea que refuerza la dinámica de poder existente, dejando a las comunidades marginadas simultáneamente en riesgo de ser removidas y sobreexpuestas a una serie de daños".[52]

Mientras que la violencia entre las personas corresponde a cuestiones culturales, políticas y sociales que requieren cambios estructurales de la misma naturaleza, la remoción de contenidos que afecta desproporcionadamente a grupos y poblaciones en situación de mayor vulnerabilidad es algo que pudiera atenderse inmediatamente por las empresas de redes sociales.

La censura por parte de las plataformas de redes sociales reduce el diálogo y el conocimiento público. Sobre todo, porque las personas periodistas, artistas, activistas y de comunidades marginadas son los principales sujetos que enfrentan la eliminación.[53]

50 ARTICLE 19, "Disonancia: voces en disputa", 26 de mayo de 2020, páginas 184-187, https://disonancia.articulo19.org/wp-content/uploads/2020/07/DISONANCIA-INF-A19-2019-PDF-WEB.pdf

51 Vilk, Viktory, *et. al.*, "No Excuse for Abuse: What Social Media Companies Can Do Now to Combat Online Harassment and Empower Users", *PEN America*, marzo de 2021, https://pen.org/report/no-excuse-for-abuse/

52 Díaz, Ángel, y Laura Hecht-Felella, *op. cit.*, página 9.

53 ARTICLE 19, "#MissingVoices. A campaign calling for better accountability and transparency", https://www.article19.org/campaigns/missingvoices/


Nisague
Nisague
Gubidxo

Consideraciones finales

Hoy en día, el uso de las redes sociales es la manera más accesible y escalable de promover causas sociopolíticas de las personas que conforman los grupos o poblaciones que son más marginados en el mundo –mismos que son los más probables de recibir abuso y censura dentro y fuera de internet–. *"[La tecnología] nos acerca a otros territorios y hace que podamos comunicarnos"*, mencionó una persona activista en defensa de la tierra y territorio, continuando *"[pero] no entiendo por qué algunas [publicaciones] se bajan y otras no"*.

Así pues, las publicaciones hechas por comunidades –como las que fueron consultadas en esta investigación– que alzan sus voces y se movilizan, y denuncian maltratos o abuso, pueden ser tomadas como discursos contrarios a las normas comunitarias de las empresas y ser eliminadas por el idioma y tipo de lenguaje que utilizan.

Hasta el momento es imposible calificar y cuantificar cuáles han sido los efectos de la remoción de contenidos en las movilizaciones de grupos y poblaciones como las que aquí fueron referidas, ya que se desconoce el número específico de eliminaciones de contenido y suspensiones de cuentas, cuántas de estas decisiones partieron de la implementación de algoritmos sin revisión humana y cuántas fueron revertidas por considerarse erróneas. Si bien existen los informes de transparencia por parte de las empresas donde éstas comparten datos relevantes, todavía hacen falta mayores niveles de granularidad de la información y más claridad en cuanto a sus acciones, en especial en lo relativo a las remociones de contenido que se generan por los sistemas automatizados.[54]

Agravando la situación, incluso sabiendo que existen este tipo de acciones de exclusión del entorno digital, aunque las plataformas de redes sociales *de jure* cuenten con sus propios procedimientos de apelación, en la práctica estos procesos no son expeditos, pertinentes y accesibles para que los daños se reparen y corrijan. *"No incumplí ninguna norma comunitaria"*, relató un activista por los derechos de la comunidad LGBTTTIQ+, *"[s]iempre soy muy cuidadoso en la construcción de mis mensajes, lo hago pensando en que los algoritmos no pueden analizar una situación específica. [...] Me puse en contacto con Twitter, [...] ni siquiera me dieron la razón por la cual me suspendieron [la cuenta]. Me llegó un mensaje diciéndome que la cuenta había sido eliminada definitivamente, que no volviera a apelar y que tampoco mandara mensaje. En Instagram también lo hice [apelar], tuve apoyo de un amigo y estuvimos mandando correos. [I]nstagram mandó un mensaje diciendo que había sido un error: me la devolvieron [la cuenta] y me la volvieron a bajar cuatro veces más. A la cuarta vez –que intenté recuperar mi cuenta– me apareció un mensaje que decía que mi cuenta ya había sido eliminada... o sea definitivamente yo ya no podía apelar. [L]as redes sociales ni siquiera siguieron el proceso [de apelación] que marcan las normas comunitarias"*.

A pesar del supuesto impulso en los últimos años para mejorar los procesos de apelación y las notificaciones de infracciones a las personas usuarias, los elementos centrales del proceso siguen siendo vagos y las opciones de apelación siguen siendo limitadas.[55] Incluso cuando se logran tramitar las apelaciones, las personas usuarias rara vez reciben una notificación adecuada sobre la norma que presuntamente violaron o el razonamiento detrás de la remoción de contenido. A menudo, las empresas identificarán el contenido infractor como uno que viola una política general amplia, como "discurso de odio", por ejemplo, sin proporcionar información suficiente y oportuna sobre qué parte de la norma fue violada.[56]

De lo explorado en el presente documento se vislumbra el gran problema que ocasiona el monopolio tecnológico en términos de derechos humanos, puesto que millones de personas están sujetas a las decisiones que concentra un puñado determinado de las plataformas digitales más populares, las cuales controlan quién tiene derecho –o no– a expresarse.[57]

Para evitar que persista la restricción ilegítima del flujo de información en internet es necesario que existan principios de gobernabilidad democrática, una infraestructura robusta, uni-

54 Consejo Asesor de Contenido, "Decisión del caso 2020-004-IG-UA", *Facebook*, 28 de enero de 2021, https://oversightboard.com/decision/IG-7THR3SI1/

55 Díaz, Ángel, y Laura Hecht-Felella, *op. cit.*, página 18.

56 *Ibidem*

57 ARTICLE 19, "#MissingVoices...", *op. cit.*

versal, accesible y cuya regulación garantice una internet libre, accesible y abierta para todas las personas usuarias.[58] Por tanto, esta discusión atañe a todos los sectores de la sociedad que tienen participación en la gobernanza de internet: las personas usuarias de las tecnologías, los gobiernos, las empresas, la sociedad civil, y la comunidad académica y técnica.

Recomendaciones

De acuerdo a la experiencia de documentación de ARTICLE 19 respecto a las remociones de contenido que sufre el gremio periodístico en las redes sociales, así como a los hallazgos provenientes de la presente investigación, se presentan las siguientes recomendaciones a las plataformas digitales respecto a sus políticas y prácticas de moderación de contenidos[59]:

- Adoptar los Principios de Santa Clara[60] sobre moderación de contenidos, con el propósito de aplicar estándares en materia de transparencia y rendición de cuentas en apego al marco internacional de los derechos humanos.

- Garantizar en cualquier contexto el derecho de acceso a la información pública, respecto de los procedimientos de remoción, moderación y eliminación de contenido. En ese sentido se deberá fortalecer el ámbito de cobertura y la granularidad de sus reportes de transparencia, de tal manera que permita someter a escrutinio público las prácticas de censura. Los reportes deben incluir con precisión y claridad los criterios y procesos que agotan para determinar la eliminación, restricción y/o desindexación de contenidos en sus plataformas, así como los mecanismos para notificar la acción o hacerla del conocimiento de terceras partes, creadoras y/o difusoras del contenido.

- Proveer procesos estandarizados de apelación ante la remoción o eliminación de contenidos o cuentas y perfiles, en aras de proporcionar una salvaguarda para la libertad de expresión en línea y permitir desafiar a estas empresas cuando tomen decisiones contrarias a derechos.

- Realizar esfuerzos para evitar que las reglas de moderación de contenidos se usen de forma abusiva para remover contenidos, ya sea en virtud de fines políticos, electorales, de controlar o restringir el flujo de información, o de socavar voces que han sido históricamente marginadas y excluidas del espacio cívico.

- Considerar los Principios de Camden sobre la Libertad de Expresión y la Igualdad al momento de formular e implementar políticas que puedan asfixiar la libre expresión, los cuales señalan que "las restricciones a la libertad de expresión apuntan a los grupos desfavorecidos y estas restricciones socavan en vez de promover la igualdad. En lugar de poner restricciones, es imprescindible permitir el debate abierto para poder combatir los estereotipos individuales y grupales negativos y para exponer el daño engendrado por la discriminación"[61].

- Aumentar la base de personas moderadoras de contenido (sin subcontrataciones o *outsourcing*) y garantizar las condiciones laborales y de seguridad necesarias[62] para que puedan desarrollar su trabajo.

- Generar las adecuaciones necesarias en los procesos de moderación de contenidos para reconocer la diversidad lingüística de los países (mejorar el filtrado de palabras, generar vínculos con organizaciones locales para convalidar expresiones locales, entre otras acciones).

- Consolidar un sistema de seguimiento de los casos de apelación que permita a las personas usuarias tener transparencia sobre la revisión de su caso –desde el inicio del reporte hasta su término–.

En cuanto a recomendaciones puntuales emitidas por personas participantes de la presente investigación, sobresalen los siguientes:

- "[Las redes sociales] *necesitan más revisores [moderadoras de contenido] adecuados a los contextos locales. Tienen que tener [en cuenta] el contexto social y económico*

58 Organización de Estados Americanos, "Declaración Conjunta del Vigésimo Aniversario: Desafíos para la Libertad de Expresión en la Próxima Década", 2019, http://www.oas.org/es/cidh/expresion/showarticle.asp?artID=1146&lID=2

59 ARTICLE 19, "#LibertadNoDisponible. Censura y remoción de contenido en Internet. Caso: México", diciembre de 2020, https://articulo19.org/wp-content/uploads/2021/02/LIBERTAD-NO-DISPONIBLE-single-page.pdf; ARTICLE 19, "Distorsión: el discurso contra la realidad", 23 de marzo de 2021, https://articulo19.org/wp-content/uploads/2021/03/Book-1_ARTICLE-19_2021_V03.pdf y ARTICLE 19, "#MissingVoices...", *op. cit.*

60 Electronic Frontier Foundation, *et. al.*, "The Santa Clara Principles On Transparency and Accountability in Content Moderation", marzo de 2020, https://santaclaraprinciples.org/

61 ARTICLE 19, "Los Principios de Camden Sobre La Libertad de Expresión y la Igualdad", abril de 2009, https://www.article19.org/data/files/pdfs/standards/los-principios-de-camden-sobre-la-libertad-de-expresion-y-la-igualdad.pdf

62 Foxglove, "Open letter from content moderators re: pandemic", 18 de noviembre de 2020, https://www.foxglove.org.uk/2020/11/18/open-letter-from-content-moderators-re-pandemic/

de la región, además del lenguaje [antes de decidir qué contenidos eliminar]", mencionó una persona activista de los derechos culturales.

- *"Entender el fondo cultural de cada región antes de implementar algoritmos sería bueno. [H]abría que trabajarse bajo un precepto de conocimientos culturales. Que participen [en la moderación de contenidos] personas antropólogas, historiadoras, sociólogas. [...] Hace falta explorar más, entendiendo el universo de expresiones que tenemos en el mundo."*, reflexionó una mujer activista por los derechos territoriales, *"Cómo te ves, cómo luces [y] cómo te expresas es algo que los algoritmos invisibilizan"*, por lo que insistió en que la moderación de contenidos necesita llevarse a cabo desde una formulación más compleja que la actual.

Si bien estas recomendaciones no resolverán tajantemente los muchos problemas existentes respecto a las vulneraciones de los derechos humanos por las redes sociales, sí son un primer paso importante para proteger –de mejor manera– la libertad de expresión, y para hacer que las empresas sean más transparentes y responsables respecto a sus acciones.[63]

63 ARTICLE 19, "#MissingVoices...", *op. cit.*

Remoción de contenidos:
Desigualdad y exclusión del espacio cívico digital

FRIEDRICH NAUMANN STIFTUNG Für die Freiheit.
México

ARTICLE 19